Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS 523, 603–619 (2023) https://doi.org/10.1093/mnras/stad1429
Advance Access publication 2023 May 11

Primordial non-Gaussianity with angular correlation function: integral constraint and validation for DES

Walter Riquelme ,[1,2★] Santiago Avila ,[1,2★] Juan García-Bellido ,[1,2★] Anna Porredon,[3,4,5] Ismael Ferrero ,[6] Kwan Chuen Chan ,[7,8] Rogerio Rosenfeld,[9] Hugo Camacho,[10,11] Adrian G. Adame,[1,2] Aurelio Carnero Rosell ,[11,12] Martin Crocce ,[13,14] Juan De Vicente ,[15] Tim Eifler ,[16,17] Jack Elvin-Poole,[3,4] Xiao Fang ,[18,19] Elisabeth Krause,[19] Martin Rodriguez Monroy,[15] Ashley J. Ross ,[3] Eusebio Sanchez[15] and Ignacio Sevilla[15]

Affiliations are listed at the end of the paper

Accepted 2023 May 4. Received 2023 May 4; in original form 2022 October 11

ABSTRACT

Local primordial non-Gaussianity (PNG) is a promising observable of the underlying physics of inflation, characterized by $f_{\rm NL}^{\rm loc}$. We present the methodology to measure $f_{\rm NL}^{\rm loc}$ from the Dark Energy Survey (DES) data using the two-point angular correlation function (ACF) with scale-dependent bias. One of the focuses of the work is the integral constraint (IC). This condition appears when estimating the mean number density of galaxies from the data and is key in obtaining unbiased $f_{\rm NL}^{\rm loc}$ constraints. The methods are analysed for two types of simulations: ∼246 GOLIAT-PNG N-body small area simulations with $f_{\rm NL}$ equal to −100 and 100, and 1952 Gaussian ICE-COLA mocks with $f_{\rm NL} = 0$ that follow the DES angular and redshift distribution. We use the ensemble of GOLIAT-PNG mocks to show the importance of the IC when measuring PNG, where we recover the fiducial values of $f_{\rm NL}$ within the 1σ when including the IC. In contrast, we found a bias of $\Delta f_{\rm NL} \sim 100$ when not including it. For a DES-like scenario, we forecast a bias of $\Delta f_{\rm NL} \sim 23$, equivalent to 1.8σ, when not using the IC for a fiducial value of $f_{\rm NL} = 100$. We use the ICE-COLA mocks to validate our analysis in a realistic DES-like set-up finding it robust to different analysis choices: best-fitting estimator, the effect of IC, BAO damping, covariance, and scale choices. We forecast a measurement of $f_{\rm NL}$ within $\sigma(f_{\rm NL}) = 31$ when using the DES-Y3 BAO sample, with the ACF in the $1\,{\rm deg} < \theta < 20\,{\rm deg}$ range.

Key words: cosmology: observations – inflation – large-scale structure of Universe.

1 INTRODUCTION

Cosmic inflation predicts that the primordial seeds, encoded in the initial gravitational potential of the Universe, are described by close to Gaussian random fields, for which all the statistical information is contained in the two-point correlation function. We can parametrize deviations from Gaussianity by using a parameter denoted by $f_{\rm NL}$, which represents the amount of primordial non-Gaussianity encoded in the three-point correlation of the fields. Primordial non-Gaussianity (PNG) is claimed to be a smoking gun to differentiate among the vast collection of inflationary models. In particular, primordial non-Gaussianity of the local type, parametrized by $f_{\rm NL}^{\rm loc}$, can distinguish between canonical single-field and non-vanilla scenarios, such as multifield inflation (Byrnes & Choi 2010; Pajer, Schmidt & Zaldarriaga 2013).

The primordial seeds affect the formation of structures at different epochs in cosmic history, implying that signals of PNG could appear in different cosmological probes. An example is the constraints of PNG coming from the cosmic microwave background (CMB) temperature bispectrum. The latest Planck results present the tightest constraints for local PNG with $f_{\rm NL}^{\rm loc} = -0.9 \pm 5.1$ (Planck Collaboration IX 2020), but since Planck reached its cosmic variance limit, another way to improve this constraint is desirable.

Similar to how PNG affects the temperature fluctuations in the CMB, the non-Gaussian initial perturbations can also affect the distribution of dark matter overdensities, which in turn affects the distribution of biased tracers of dark matter (e.g. galaxies, quasars). This implies that PNG could also be constrained using the bispectrum of such tracers, as has been studied in Jeong & Komatsu (2009), Tasinato et al. (2014), and Moradinezhad Dizgah et al. (2021).

Given the complexity of modelling the bispectrum, dominated by late non-Gaussianities induced by non-linear evolution,[1] and other difficulties such as non-linear bias, redshift space distortions, and the window function of the survey (Gil-Marín et al. 2017; Sugiyama et al. 2019), a different method to look for primordial non-Gaussianity using late-time objects is desired. Another effect of PNG is on the halo formation mechanism. Local primordial non-Gaussianity induces a scale dependence on the linear bias between galaxies and the underlying dark matter overdensities. The scale dependence in

[1] It is worth mentioning that besides these difficulties, recent work using the EFT of LSS for the bispectrum has proven to be helpful when constraining local PNG from eBOSS data (Cabass et al. 2022).

★ E-mail: walter.riquelme@uam.es (WR); santiago.avila@uam.es (SA); juan.garciabellido@uam.es (JG-B)

the bias creates a characteristic signal in the two-point correlation at very large scales, which can be constrained using different large-scale structure (LSS) biased tracers (Dalal et al. 2008; Matarrese & Verde 2008; Slosar et al. 2008). Some studies show that PNG can also be constrained using galaxies with zero linear bias in low-density environments (Castorina et al. 2018), or even negatively biased traces, such as voids (Chan, Hamaus & Biagetti 2019).

Measurements of cosmological parameters using two-point correlation functions have been done multiple times because they are easy to model and have a large signal-to-noise ratio. This makes the scale-dependent bias in the two-point correlation the more robust method to constrain PNG. Previous measurements of PNG using the scale-dependent bias have been presented in Slosar et al. (2008), Ross et al. (2013), Giannantonio et al. (2014), Leistedt, Peiris & Roth (2014), Ho et al. (2015), Castorina et al. (2019), and Mueller et al. (2021).

One noticeable trend is that most of the current constraints come from spectroscopic surveys. It has been shown in de Putter & Doré (2017) that imaging surveys with high volumes could overcome redshift uncertainties and had the potential of breaking the $\sigma(f_{\rm NL}^{\rm loc}) \sim 1$ barrier. Hence, upcoming photometric data from the Legacy Survey of Space and Time (LSST) in the Vera Rubin Observatory[2] (LSST Science Collaboration 2009) is a promising source to break current bounds.

This work is a first step to measure PNG with existing data from the Dark Energy Survey (DES[3]; DES Collaboration 2021), which represents the state of the art in photometric surveys. Currently, the DES has surveyed over ∼388 million galaxies in ∼5000 deg^2 and presents an opportunity to put the tightest constraints from photometric surveys (as will see in this work).

DES has successfully probed the nature of dark energy using different cosmological probes (DES Collaboration 2018; DES Collaboration 2022a; Porredon et al. 2021; Rodríguez-Monroy et al. 2022). One of them is the study of clustering of galaxies for the measurement of the Baryon Acoustic Oscillation (BAO) scale (DES Collaboration 2019, DES Collaboration 2022b) using galaxy data. The BAO scale measurement suggests that we could also use clustering of galaxies at large scales for measuring PNG within DES.

This work presents the starting point in this direction by describing the methods to constrain the $f_{\rm NL}^{\rm loc}$ parameter using DES simulations. We use the angular correlation function (ACF) as a summary statistic for the galaxy distribution and show the effect that primordial non-Gaussianities have on the angular clustering of galaxies via the scale-dependent bias.

One of the main focuses of the work is on the integral constraint (IC; Groth & Peebles 1977; Peacock & Nicholson 1991; Ross et al. 2013; Beutler et al. 2014; de Mattia & Ruhlmann-Kleider 2019). The IC corrects the modelled correlation function by adding a constant, which comes from imposing that its integral over the whole survey volume needs to vanish. This correction is found to be key to obtaining unbiased PNG measurements.

The IC was not relevant in the previous DES non-PNG clustering analysis for two main reasons: First, its effect becomes relevant at very large scales. Secondly, for the case of BAO measurements, its template includes marginalization over nuisance parameters, one of them being a constant shift in the amplitude of the ACF. This shift mimics the IC correction, implying that any effect from it has already been marginalized.

[2] https://www.lsst.org/
[3] https://www.darkenergysurvey.org/

In this paper, we use the ACF with PNG, and the IC, as a theoretical template to measure the value of $f_{\rm NL}^{\rm loc}$ from simulated galaxy catalogues. The measurement is based on Bayesian parameter inference using MCMC (Markov chain Monte Carlo) sampling of a Gaussian likelihood function. The methods are analysed for two kinds of simulations. First, we introduce the GOLIAT-PNG mocks (Avila & Adame 2023), a set of 246 *N*-Body simulations that have non-Gaussian initial conditions. We use these simulations to remark on the importance of the IC when measuring $f_{\rm NL}^{\rm loc}$. Secondly, we use 1952 ICE-COLA mocks (Ferrero et al. 2021) that follow the DES angular and redshift distribution of the Y3 BAO galaxy sample (Carnero Rosell et al. 2022) to validate the pipeline. We show that it is robust against different analysis choices, such as covariance modelling, $f_{\rm NL}$ estimator, and scale cuts. Finally, we forecast a measurement of the accuracy of $f_{\rm NL}^{\rm loc}$ when using the DES Y3 BAO sample data.

This paper is organized as follows. The steps to model ACF with scale-dependent bias are presented in Section 2. In Section 3, we derive the IC and show its importance when dealing with local PNG. In Section 4, we describe the simulations that we will use to test and optimize the methods. Section 5 presents the tools needed to extract the $f_{\rm NL}$ parameter. In Section 6, we test the pipeline against the GOLIAT-PNG simulations and show how the IC is needed to obtain unbiased values of $f_{\rm NL}$. Once the methods are tested over non-Gaussian simulations, we validate the pipeline using ICE-COLA simulations in Section 7.

2 THEORY

In this section, we describe the impact of PNG on the two-point statistics of biased tracers. First, we describe how non-Gaussian initial conditions modify the bias relation, introducing the scale-dependent bias. After, we show the effect that it has on the power spectrum. Finally, we focus on the ACF and show how it is affected by local primordial non-Gaussianity.

2.1 Gaussian galaxy bias

The spatial distribution of matter is set by the initial conditions coming from cosmic inflation, which predicts a nearly scale-invariant power spectrum and a close to Gaussian distribution for the primordial gravitational fields. During the matter domination era, dark matter collapsed due to these gravitational potentials generating haloes which, as the Universe evolves, will serve as the backbones for the creation of large-scale structures.

We will focus our analysis on angular separations of galaxies larger than 1 deg. This choice is customary for the BAO analysis because such scales are within the linear regime of perturbation theory, simplifying the theoretical modelling (DES Collaboration 2022b). In this regime, galaxies follow the trace of the dark matter overdensities by the linear relation,

$$\delta_{\rm g}(\boldsymbol{x}) = b\,\delta_{\rm m}(\boldsymbol{x}), \tag{1}$$

where b is a parameter called galaxy bias, which is found to be constant at large scales under the standard Gaussian initial conditions.

In the non-linear regime, non-linear effects also generate a scale-dependent bias, which affects only small scales. We will ignore such effects throughout this work and refer the reader to Desjacques, Jeong & Schmidt (2018) for an intensive review on the scale dependence of the galaxy bias and other related effects.

The statistical distribution of dark matter overdensities is well described by the matter power spectrum $P_{\rm m}(k)$, which depends on the primordial power spectrum, coming from inflation, and the transfer function $T(k)$, which describes its evolution throughout cosmic history. Due to equation (1), the biased relation between galaxies and dark matter also appears in the galaxy power spectrum as follows:

$$P_{\rm g}(k) = b^2 P_{\rm m}(k). \tag{2}$$

As we will see in the following section, the linear relation between galaxies and dark matter will change when dealing with non-Gaussian initial conditions.

2.2 PNG via scale-dependent bias

Deviations from Gaussianity in the initial conditions, coming from inflation, is an active area of research due to the potential of unveiling the nature of the primordial fields. In particular, we focus on PNG of the local type (Komatsu & Spergel 2001),

$$\Phi_{\rm NG}(\mathbf{x}) = \phi_{\rm G}(\mathbf{x}) + f_{\rm NL}^{\rm loc}(\phi_{\rm G}^2(\mathbf{x}) - \langle\phi_{\rm G}^2\rangle), \tag{3}$$

where $\Phi_{\rm NG}(\boldsymbol{x})$ is the non-Gaussian Newtonian potential and $\phi_G(\boldsymbol{x})$ is the Gaussian potential. Under this approximation, $f_{\rm NL}^{\rm loc}$ is a constant that parametrizes deviations from Gaussian initial conditions. Throughout this work, we will focus on local PNG; hence, from here on, we will drop the superscript 'loc' for simplicity.

Dalal et al. (2008) and Slosar et al. (2008) showed that PNG, parametrized as equation (3), would change the way dark matter collapses into halos, subsequently affecting galaxy formation. In the presence of local PNG, the long wavelength modes of the primordial gravitational potential couple with the smaller modes, responsible for the local amplitude of matter fluctuations, producing a modulation in the local number density of haloes. The change in the local number density will add an extra contribution to the galaxy bias, which depends on the scale. We can write the scale-dependent bias due to local PNG as follows:

$$b(k) = b + f_{\rm NL}\alpha(k, z)\frac{\partial \ln n}{\partial \ln \sigma_8}, \tag{4}$$

where b is the constant linear bias and $\delta_c = 1.686$ is the critical value of collapse for halo formation in an Einstein-de Sitter universe (Fillmore & Goldreich 1984). Also,

$$\alpha(z, k) = \frac{3\Omega_m}{2D(z)}\frac{H_0^2}{c^2}\frac{g(0)}{g(z_{\rm rad})}\frac{1}{k^2 T(k)}, \tag{5}$$

where H_0 is the Hubble factor today,[4] c is the light speed, and $\Omega_{\rm m}$ is the matter density today. In addition, $T(k)$ is the linear transfer function, and $D(z)$ is the linear growth factor, both normalized to 1 at $k = 0$ and $z = 0$, respectively. The factor $\frac{g(0)}{g(z_{\rm rad})}$, with $g(z) = (1 + z)D(z)$, arises because $D(z)$ is normalized to unity and can be omitted if normalized to the scale factor during the matter-dominated era (Mueller, Percival & Ruggeri 2019). Its value is shown to be $\frac{g(0)}{g(z_{\rm rad})} \simeq 1.3$.[5]

One particularity of this scale-dependent bias is its $1/k^2$ dependence, implying that primordial non-Gaussianity affects the distribution of galaxies only at very large scales. Throughout this work, we will refer to scale-dependent bias as the one produced due to primordial non-Gaussianity.

It has been shown in Slosar et al. (2008) that

$$\frac{\partial \ln n}{\partial \ln \sigma_8} = 2\delta_c(b - p), \tag{6}$$

where the parameter p was introduced to show deviations from the original model of Dalal et al. (2008) to take into account different tracers. We refer the reader to Barreira (2020) for an analysis of the impact of the parameter p and other assumptions on the non-Gaussian bias. For the case of ICE-COLA mocks, we will fix $p = 1$, which is customary in many analyses and is considered the prediction for a mass-selected galaxy/halo sample. Finally, the scale-dependent bias we will use in this work can be written as follows:

$$b(k) = b + 2(b - p) f_{\rm NL}\alpha(k, z)\delta_c. \tag{7}$$

As an example of the effect of the scale-dependent bias, in Fig. 1, we compute the linear matter power spectrum from CAMB[6] (Lewis, Challinor & Lasenby 2000; Howlett et al. 2012) and apply a scale-dependent bias as given in equation (7) to show the galaxy power spectrum for different values of $f_{\rm NL}$. The power spectrum is computed using the cosmological parameters from the ICE-COLA simulation presented in Section 4.2.

Since the scale-dependent bias is squared in the galaxy power spectrum, we will have contributions with different dependence on $f_{\rm NL}$. This dependence can be seen as follows:

$$b(k)^2 \propto b^2 + A\, b\frac{f_{\rm NL}}{k^2} + B\frac{f_{\rm NL}^2}{k^4}, \tag{8}$$

where A and B are prefactors that do not depend on the scale [since $T(k)$ becomes constant at very large scales]. The previous equation tells us that we have quadratic and linear terms in $f_{\rm NL}$ and a term that does not depend on $f_{\rm NL}$. Fig. 1 shows how the scale-dependent bias generates an enhancement of the power spectrum at large scales for $f_{\rm NL} = 100$. The situation is more interesting for $f_{\rm NL} = -100$, where the linear term in $f_{\rm NL}$ generates a reduction in the power spectrum until a given scale, then the quadratic term overcomes, explaining the sharp feature around at $k = 0.005\, h\,{\rm Mpc}^{-1}$.

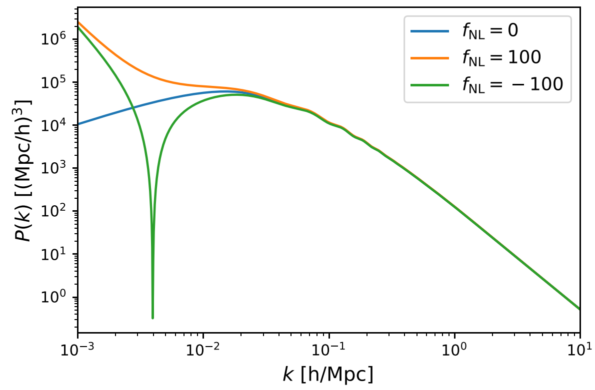

Figure 1. Theoretical linear galaxy power spectrum with scale-dependent bias for $f_{\rm NL} = 0$ (blue line), $f_{\rm NL} = 100$ (orange line), and $f_{\rm NL} = -100$ (green line). The power spectrum is computed using the fiducial cosmological parameters of the GOLIAT-PNG simulations described in Table 1.

[4] If one uses k in units of hMpc^{-1}, then $H_0 = 100h$ [Mpc^{-1} km s^{-1}] with $h = 0.7$.

[5] This value is slightly cosmology dependent. When comparing against the ICE-COLA mocks, we will consider it as 1.3 since we do not expect that it affects the constraints if we plan to recover $f_{\rm NL} = 0$. On the other side, for the non-Gaussian GOLIAT-PNG simulations, it was shown to be 1.316 for the fiducial cosmology of the simulations.

[6] https://camb.readthedocs.io

2.3 BAO-damped galaxy power spectrum

We may need to use precise theoretical modelling to obtain an optimal measurement of $f_{\rm NL}$. For this, we follow the methodology used in DES Collaboration (2022b) for the DES Y3 BAO template, based on extensions of the linear power spectrum using IR resummation methods optimized for an accurate description of the damping in the BAO peak (Blas et al. 2016; Ivanov & Sibiryakov 2018). The particularity of this method relies on a derivation of the BAO damping based on first principles, in contrast with other models where the damping is obtained from fits over simulations. In Section 7.3, we will compare the impact of using the BAO-damped galaxy power spectrum versus linear theory without damping on the $f_{\rm NL}$ measurement.

The BAO-damped galaxy power spectrum is given by

$$P(k,\mu,z) = (b(k) + f(z)\mu^2)^2\,[(P_{\rm lin}(k) - P_{\rm nw}(k))D_{\rm BAO} + P_{\rm nw}(k)]\,, \tag{9}$$

where $P_{\rm lin}(k)$ is the linear matter power spectrum. $P_{\rm nw}(k)$ is the smooth 'no-wiggle' power spectrum. We refer the reader to DES Collaboration (2022b) for further details on how to compute it. The function $f(z)$ is the growth rate of structures, defined under the following approximation (Linder 2005),

$$f(z) \approx \Omega_m(z)^\gamma, \tag{10}$$

with $\gamma = 0.55$. The parameter μ is defined as the cosine of the angle between the line of sight and wave vector $\boldsymbol{k}$.

In equation (9), $D_{\rm BAO}(z)$ is a Gaussian damping defined by

$$D_{\rm BAO}(z) = \exp\{-k^2(\mu^2\Sigma_\parallel^2 + (1-\mu^2)\Sigma_\perp^2 + f(z)\mu^2(\mu^2-1)\delta\Sigma^2)\}, \tag{11}$$

where $\Sigma_\parallel(z) = (1 + f(z)\Sigma_\perp)$. The parameters $\Sigma_\perp$ and $\delta\Sigma$ can be computed directly for a fixed cosmology. In the case of ICE-COLA cosmology, at $z = 0$, $\Sigma_\perp = 5.8$ Mpc h^{-1} and $\delta\Sigma = 3.18$ Mpc h^{-1} and they are scaled by the growth factor to any other redshift (DES Collaboration 2022b).

When comparing against the ICE-COLA simulations, we will include the BAO damping in the power spectrum, as presented in this subsection, since we will be using these simulations to validate the methods and improve the accuracy for $f_{\rm NL}$, implying the need for a more precise theory modelling. When comparing against GOLIAT-PNG simulations, we will not consider BAO damping because we use those simulations to recover higher $f_{\rm NL}$ values, and we do not expect the damping to be a determinant factor in their accuracy. We will come back to this discussion on Section 7.3, where we will assess the impact of the BAO damping on the $f_{\rm NL}$ measurement. Also, notice that the scale-dependent bias described in the previous subsection is already added in equation (9), adding extra contributions to the galaxy power spectrum.

With the previously computed power spectrum, we can use a multipole expansion in Legendre polynomials of μ,

$$P_\ell(k,z) \equiv \frac{(2\ell+1)}{2}\int_{-1}^{1} {\rm d}\mu\, P(k,\mu,z)L_\ell(\mu), \tag{12}$$

to take into account the anisotropies caused by redshift space distortions to the line of sight. Notice that the power spectrum is computed at $z = 0$ and does not include the growth factor $D(z)$ since this will be added when calculating the ACF in the next section.

2.4 ACF with PNG

Using the previously described power spectrum, we can compute its configuration space counterpart, the two-point correlation function (2PCF), using the multipole expansion of equation (12),

$$\xi(r, \hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{l}}) = \sum_{\ell=0,2,4} \xi_\ell(r)L_\ell(\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{l}}), \tag{13}$$

$$\xi_\ell(r) = \frac{i^\ell}{2\pi^2}\int_0^\infty {\rm d}k\, k^2 j_\ell(kr)P_\ell(k,\bar{z}), \tag{14}$$

where r is the separation distance between galaxies and j_ℓ is the spherical Bessel function. Notice that the previously computed power spectrum is evaluated at the mean redshift of the photo-z distribution, $\bar{z}$. The correlation function is also a function of the angle between the line of sight direction $\hat{\boldsymbol{l}}$ and the direction of the separation vector $\hat{\boldsymbol{r}}$, given by

$$\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{l}} = \frac{\chi(z_2)-\chi(z_1)}{r}\cos\frac{\theta}{2}, \tag{15}$$

where $\chi(z)$ is the comoving distance, and θ is the angular separation between two galaxies.

It is important to remember that because of primordial non-Gaussianity, we now have a scale-dependent bias $b(k)$ that will be a part of each $P_\ell(k, z)$ and needs to be considered for the computation of the 2PCF.

We can compute the ACF (Crocce, Cabré & Gaztañaga 2011a; Chan et al. 2018) as the two-dimensional projection of the 2PCF following the galaxy photo-z distribution, $N(z)$, normalized such that its integral over redshift is equal to 1. With this, the ACF is given by

$$w(\theta) = \int {\rm d}z_1 \int {\rm d}z_2\phi(z_1)\phi(z_2)\xi(r(z_1,z_2,\theta),\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{l}}), \tag{16}$$

which is a function of the angular separation defined through the relation,

$$r(z_1,z_2,\theta) = \left(\chi(z_1)^2 + \chi(z_2)^2 - 2\chi(z_1)\chi(z_2)\cos\theta\right)^{1/2}. \tag{17}$$

where $\phi(z) = N(z)D(z)$. The previously obtained power spectrum was computed at $z = 0$, so $\phi(z)$ incorporates its evolution to a different redshift.

As mentioned before, the theoretical ACF with PNG shares similarities with the BAO template, but adding extra terms proportional to $f_{\rm NL}$, to clarify this, we can consider that our PNG template is composed of a BAO-part and a $f_{\rm NL}$-part, as follows:

$$w(\theta) = w_{\rm BAO}(\theta) + w(\theta, f_{\rm NL}), \tag{18}$$

where $w_{\rm BAO}(\theta)$ is the BAO template used in DES Collaboration (2022b), schematically given by

$$w_{\rm BAO}(\theta) \sim b^2 w_b(\theta) + bf\,w_{bf}(\theta) + f^2 w_f(\theta), \tag{19}$$

where $w_{b,bf,f}(\theta)$ correspond to different ACF contributions arranged by their pre-factors. On the other hand, the $f_{\rm NL}$-part involves the extra terms proportional to $f_{\rm NL}$, in accordance with equation (8), as follows:

$$w(\theta, f_{\rm NL}) \sim bf_{\rm NL}w_{\rm A}(\theta) + f_{\rm NL}^2 w_{\rm B}(\theta) \tag{20}$$

where $w_{\rm A,\,B}(\theta)$ involve the scale-dependent contributions of the ACF. As a reminder of this discussion, we will extend the notation of our theoretical modelling to

$$w(\theta) \rightarrow w_{\rm th}(\theta, f_{\rm NL}), \tag{21}$$

highlighting its dependence on $f_{\rm NL}$.

The behaviour of the angular correlation with PNG can be seen in Fig. 2, where we compute the ACF using the BAO damped power

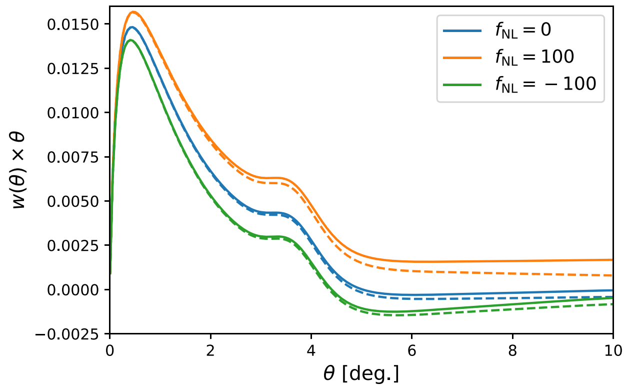

Figure 2. Theoretical ACF with the scale-dependent bias for $f_{\rm NL} = 0$ (blue line), $f_{\rm NL} = 100$ (orange line), and $f_{\rm NL} = -100$ (green line) for the first redshift bin using the ICE-COLA configuration as presented in Table 1. The solid lines are without IC. The dashed lines are with the IC correction, as discussed in Section 3, computed using equation (39) with the ICE-COLA angular footprint.

spectrum, with linear bias and $N(z)$ from the first redshift bin of the ICE-COLA mocks. As expected, we show that primordial non-Gaussianity induces a large-scale enhancement of clustering in the ACF of galaxies due to the scale-dependent bias. It can be noticed that the sharp feature in the power spectrum for $f_{\rm NL} = -100$, produced due to the linear term in $f_{\rm NL}$ (equation 8), has now translated into a small overall rising at scales around ∼10 deg (solid green line in Fig. 2). This rising is due to the integration of the Fourier transform to compute the 2PCF. As a preview of the upcoming section, we also show the IC's effect on the theoretical model. The main discussion of the upcoming section will be on how to compute the IC correction and the effect on the ACF.

3 IC AND $f_{\rm NL}$

In this section, we comment on how the excess of clustering at large scales, due to scale-dependent bias, on the theoretical angular correlation is suppressed by imposing that its integral over the survey volume needs to vanish. This condition is known as the IC.

We discuss how the IC arises from an observational point of view. We also remark on its dependence on $f_{\rm NL}$ and show how to correct the theoretical template to incorporate its effect.

3.1 Observational IC

3.1.1 IC from the observed 2PCF

Let us start with the statistical definition of the two-point correlation function for galaxies $\xi_{\rm obs}(r)$,

$$\mathrm{d}P = \bar{n}(1 + \xi_{\rm obs}(r))\mathrm{d}V, \tag{22}$$

where P is the probability of finding two objects within the volume V separated by a distance r (Peebles 1980) and $\bar{n}$ is the mean number density of galaxies in the Universe. If we integrate equation(22) over the volume of a survey, we find out that

$$N_g = \bar{n} \int \mathrm{d}V_s + \bar{n} \int \xi_{\rm obs}(r)\mathrm{d}V_s, \tag{23}$$

where N_g is the expected number of galaxies within the survey region and V_s is the total volume of the survey. Since the expected number of galaxies within the survey volume is chosen to be obtained from the survey mean number density, we have the following:

$$N_g = \bar{n} \int \mathrm{d}V_s. \tag{24}$$

The previous equation implies a condition that needs to hold for the observed two-point correlation function of galaxies within the survey volume,

$$\int \xi_{\rm obs}(r)\mathrm{d}V_s = 0. \tag{25}$$

This is the IC condition. We can re-write the IC condition as follows:

$$\int \xi_{\rm obs}(\boldsymbol{r})\mathrm{d}V_s = \int \mathrm{d}^3\boldsymbol{r} \int \mathrm{d}^3\boldsymbol{r}_1 W(\boldsymbol{r}_1)W(\boldsymbol{r}_1 - \boldsymbol{r})\xi(\boldsymbol{r}) = 0, \tag{26}$$

where $W(\boldsymbol{r})$ is the selection function for a volume-limited survey and $\boldsymbol{r} = \boldsymbol{r}_1 - \boldsymbol{r}_2$.

The previous expression can be computed directly for a given survey selection function. The problem is that defining the volume of a survey is a difficult task. Instead, it is most common to construct a random catalogue of galaxies following the shape of the survey mask to model the survey volume as pair counts between the random catalogues.

As previously mentioned, the number count of galaxies within a homogeneous region can be computed as a volume integral of the selection function,

$$N_g = \bar{n} \int \mathrm{d}^3\boldsymbol{r}\, W(\boldsymbol{r}). \tag{27}$$

Therefore, the number of random-random pair correlations, $RR(\boldsymbol{r})$, can be computed as the correlation of the number of random objects within the limited region [see e.g. Breton & de la Torre (2021), and references therein],

$$RR(\boldsymbol{r}) = \langle N_1 N_2 \rangle = \bar{n}^2 \int \mathrm{d}^3\boldsymbol{r}_1 W(\boldsymbol{r}_1)W(\boldsymbol{r}_1 - \boldsymbol{r}), \tag{28}$$

with $\boldsymbol{r} = \boldsymbol{r}_1 - \boldsymbol{r}_2$. Using the previous equation, we can compute the volume integral over a window function, and inserting equation (28) into equation (26), we obtain the following:

$$\int \xi_{\rm obs}(\boldsymbol{r})\mathrm{d}V_s = \frac{1}{\bar{n}^2} \sum_{\text{all pairs}} RR(\boldsymbol{r})\xi(\boldsymbol{r}), \tag{29}$$

where now we sum over all the possible separations between galaxies within a limited survey size. This implies that the IC condition, equation (25), can be written in terms of the $RR(\boldsymbol{r})$ pairs, as follows:

$$\sum_{\text{all pairs}} RR(\boldsymbol{r})\xi(\boldsymbol{r}) = 0, \tag{30}$$

where, for simplicity, the random–random pair correlations can be obtained from random catalogues that follow the survey mask instead of using the analytic expression.

3.1.2 IC in the observed ACF

The previous procedure can be extended to the ACF. The starting point is now the probability of finding two galaxies in a two-dimensional projection of the sky separated by an angular separation θ, as follows:

$$\mathrm{d}P = \bar{n}(1 + w_{\rm obs}(\theta))\mathrm{d}\Omega, \tag{31}$$

where $w_{\rm obs}(\theta)$ is the observed ACF.

This implies that the IC can be extended to the ACF in the same way as equation (29),

$$\int \mathrm{d}\Omega_1 \int \mathrm{d}\Omega_2 W(\hat{\boldsymbol{r}}_1)W(\hat{\boldsymbol{r}}_2)w_{\rm obs}(\theta) = 0, \tag{32}$$

where $W(\hat{\boldsymbol{r}})$ is the angular selection function, and θ is the angle subtended by $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$.

The calculation of the volume integral in the previous subsection can be extended to the sum of random-random angular pairs. This implies that we can compute the IC for the angular correlation as follows:

$$\sum_{\Omega} RR(\theta) w_{\text{obs}}(\theta) = 0, \tag{33}$$

where now the sum is over all the possible angular separations allowed by the survey mask. Also, as before, the random-random pairs correlation is obtained from the random catalogues. In practice, since we have a $w(\theta)$ for each redshift bin, this condition is applied to each of them individually.

3.2 Theoretical IC

Up to this point, we have only presented a condition that the correlation function needs to accomplish in limited surveys, and they certainly do for the usual observed correlation functions. A problem arises when we compare the theory with PNG to observational data.

3.2.1 Gaussian case

Let us start from a theoretical point of view without considering PNG. The matter power spectrum at large scales exhibits behaviour that goes as

$$P_{\text{m}}(k) \propto k^{n_s}, \tag{34}$$

where n_s is close to 1. This implies that the matter power spectrum vanishes when $k \to 0$, and since the power spectrum is related to the variance of the overdensities, this is an insight that the matter fluctuations of our Universe reach homogeneity at very large scales.

The vanishing of the matter power spectrum at large scales implies a condition to its configuration space counterpart, the 2PCF, which can be seen as follows:

$$P_m(k=0) = \int \xi(\boldsymbol{r}) \text{d}^3\boldsymbol{r} = 4\pi \int_0^{\infty} \xi(r) r^2 \text{d}r = 0. \tag{35}$$

This is the IC condition presented in equation (26) but now coming from a purely theoretical perspective.

Without the effect of f_{NL}, this same condition is expected to hold for the linear galaxy power spectrum since a linear bias relates both power spectra, and there is no change in the shape of the power spectrum. Hence, in the case of an ideal homogeneous infinite survey, the theoretical model already satisfies the observational IC. When the effect of the window function becomes more pronounced (due to either strong inhomogeneities in the randoms or small explored volumes), we will need to adjust the theory to fulfil the IC condition (see Section 3.3).

3.2.2 IC in the presence of PNG

The situation now changes in the presence of PNG. The scale-dependent bias between the galaxies and matter overdensities will modify the shape of the galaxy power spectrum introducing a $1/k^2$ correction to the matter power spectrum that depends on f_{NL}, as described in equation (5). The scale-dependent bias will generate an enhancement of the galaxy power spectrum at large scales ($k \ll k_{\text{eq}}$) with the following divergent behaviour:

$$P_g(k \to 0, f_{\text{NL}}) \sim \left(f_{\text{NL}} \frac{b-1}{k^2} \right)^2 \cdot k^{n_s} \to \infty \tag{36}$$

where $P_g(k)$ is the galaxy power spectrum. This divergence that the volume integral over the 2PCF (equation 35) will diverge for this case. As a side note, since for our modelling, we integrate numerically, the previously mentioned divergence will turn into a large (but finite) number that could depend on the integration method or resolution. Since the ACF is an integral of the 3D 2PCF (equation 16), $w(\theta)$ will have a divergence proportional to f_{NL}^2.

The discussion of this section tells us that, even if we have an infinite homogeneous survey with a negligible window function effect, the IC condition will not be fulfilled for the case of $f_{\text{NL}} \neq 0$. Additionally, the theoretical model will contain an arbitrary additive constant that depends on f_{NL}^2. This dependence will bias any results when using this model to constrain f_{NL}. This remarks the importance of the IC condition when dealing with PNG, implying that we need correct our modelling to consider this issue.

As a verification of the issue, in appendix A, we show an analytical example that illustrates how the IC condition looks for a simplified theoretical two-point correlation function in the presence of PNG. We show explicitly that the integral of the 2PCF diverges at large scales and is proportional to f_{NL}^2, implying that imposing the observational IC condition is very important when dealing with PNG analysis.

3.3 IC correction

To surpass the problem described in the previous subsection, we define an IC-corrected theoretical ACF,

$$w^{\text{IC}}(\theta, f_{\text{NL}}) = w_{\text{th}}(\theta, f_{\text{NL}}) - I(f_{\text{NL}}), \tag{37}$$

where $I(f_{\text{NL}})$ parametrize deviations from the observed IC condition (equation 33) as follows:

$$\sum_{\Omega} RR(\theta) w^{\text{IC}}(\theta, f_{\text{NL}}) = 0. \tag{38}$$

This implies that the IC correction, $I(f_{\text{NL}})$, is given by

$$I(f_{\text{NL}}) = \frac{\sum^{\theta_{\text{lim}}} RR(\theta) w_{\text{th}}(\theta, f_{\text{NL}})}{\sum^{\theta_{\text{lim}}} RR(\theta)}. \tag{39}$$

where θ_{lim} is the maximum limit angular separation allowed for the angular survey window. The effect of the IC in the context of PNG has been previously addressed in Ross et al. (2013) and Mueller et al. (2021) for the power spectrum and in Ross et al. (2013) for the 2PCF. The novelty of this work is to present a detailed analysis of its effect on the ACF and show its importance when dealing with PNG simulations, as we show in Section 6.

4 SIMULATIONS

In this section, we present the simulations that we used for testing the theoretical modelling and the validation of the f_{NL} measurements.

4.1 GOLIAT-PNG

In order to test our analysis pipeline, we first consider the use of simulations with Primordial non-Gaussianity included. Whereas many tests can be done with Gaussian initial conditions (see Section 7), there are validation steps that require PNG mocks to show the validity of the pipeline. In particular, in this work, only when fitting PNG mocks can we realize the paramount importance of including the IC.

The GOLIAT-PNG suite (Avila & Adame 2023) consists of a series of N-body simulations with ΛCDM + local PNG cosmology with $\Omega_{\text{m}} = 0.27$, $\Omega_{\text{b}} = 0.044$, $h = 0.7$, $n_{\text{s}} = 0.96$, $\sigma_8 = 0.8$, and three values for PNG: $f_{\text{NL}} = -100, 0, +100$. A summary of the

Table 1. Summary of the fixed cosmological parameters and the free measured parameters with the priors considered. The squared brackets represent flat priors.

	GOLIAT-PNG	
Parameter	Fiducial	Prior
Ω_m	0.27	–
Ω_Λ	0.73	–
Ω_b	0.044	–
n_s	0.96	–
σ_8	0.8	–
h	0.7	–
f_{NL}	−100, 100	[−700, 700]
Linear bias b	2.35	[1, 3]
Integral constraint I_i	–	[− 0.1, 0.1]
Footprint area (deg^2)	396.06	
z_{mean}	1	–
	ICE-COLA	
Parameter	Fiducial	Prior
Ω_m	0.25	–
Ω_Λ	0.75	–
Ω_b	0.044	–
n_s	0.95	–
σ_8	0.8	–
h	0.7	–
f_{NL}	0	[−500, 500]
Linear bias b_i	1.60, 1.60, 1.68, 1.82, 2.02	[1, 3]
Integral constraint I_i	–	[−0.1, 0.1]
Footprint area (deg^2)	4108.47	
z_{mean}	0.65, 0.74, 0.84, 0.94, 1.02	–

cosmological parameters and fiducial values used is presented in the first part of Table 1. The simulations have a box size of $L = 1$ Gpc h^{-1}. The initial conditions are set at $z = 32$ with second-order Lagrangian perturbation theory (2LPT) using the public code 2LPTIC[7] and evolved to $z = 1$ with GADGET2.[8] Subsequently, the $z = 1$ dark matter snapshots are run through the Amiga Halo Finder[9] to construct the halo catalogues with a minimum of 10 particles, which yield $M_h \sim 5 \times 10^{12}$ $M_\odot$ as the halo mass resolution.

Also, for the GOLIAT-PNG simulations, it was found that $p = 0.90$ for $f_{NL} = 100$ (Avila & Adame 2023), and $p = 0.92$ for $f_{NL} = -100$, when measuring their real space power spectra, and we will consider this when measuring f_{NL} from these mocks.

Another particularity of these simulations is that the initial conditions are run with the fixed and paired initial conditions (Angulo & Pontzen 2016) aimed at reducing the sample variance of the ensemble average of the two-point functions measured from these simulations. In the context of PNG, this technique is validated in Avila & Adame (2023), and we refer the reader there for further details of the GOLIAT-PNG suite. We use 41 pairs of simulations for each value of f_{NL}.

Finally, we transform those mocks from the cubic box into observable coordinates {RA, Dec., z} by setting an observer 1556 Mpc h^{-1} away from the centre of one of the faces of the box. This transformation allows us to have a mock survey with a circular semi-aperture of 11.2 deg, covering an area of roughly 396 deg^2, and a redshift range of $0.6 < z < 1.1$, the shape and size of the mask can be seen

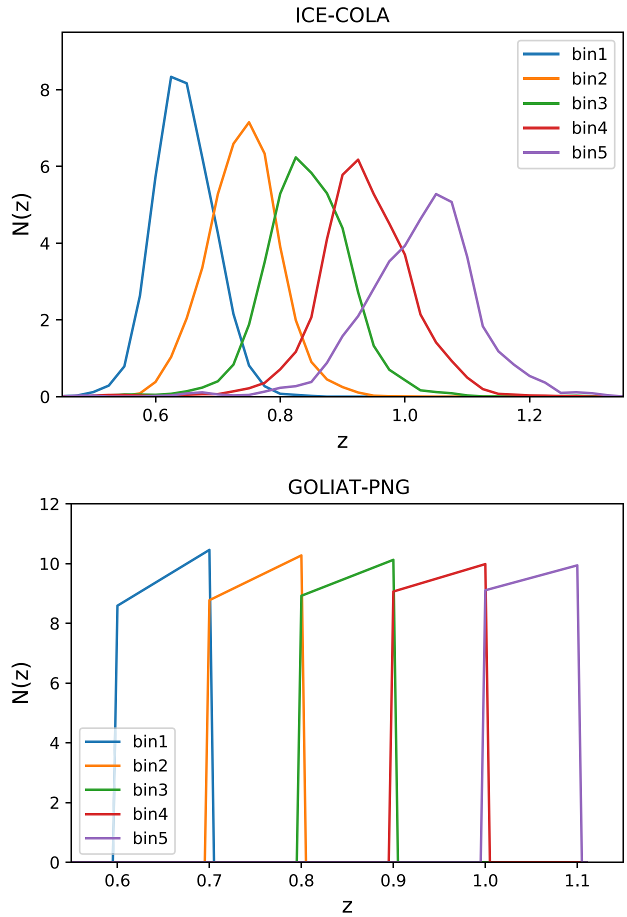

Figure 3. $N(z)$ distribution as a function of redshift for each redshift bin for the ICE-COLA mocks (top) and for the GOLIAT-PNG mocks (bottom). We remark that the GOLIAT-PNG simulations do not have photo-z errors included, implying that they do not represent a realistic $N(z)$ distribution, in contrast to the ICE-COLA mocks.

in Fig. 4. We further split the mocks into five redshift bins between 0.6 and 1.1 with $\Delta z = 0.1$. This, together with a constant number density of haloes, gives the redshift distribution $N(z)$ shown in Fig. 3. However, we note that we do not introduce any redshift space distortions, redshift error, HOD model, or even temporal evolution. We built everything from the halo catalogue at the comoving output at a redshift of $z = 1$ and a fixed halo mass threshold. This implies that we fix $D(z = 1)$ in equation (16) when using the GOLIAT-PNG mocks. We also consider three different rotations (one per Cartesian axis) for constructing the mocks, eventually resulting in a total of 246 mocks for each value of f_{NL}.

4.2 ICE-COLA

The ICE-COLA mocks (Ferrero et al. 2021) are the second set of simulations we count on for analysing and validating our methods. This set of 1952 mock galaxy catalogues is designed to mimic the DES Year 3 BAO sample (Carnero Rosell et al. 2022) over its full photometric redshift range $0.6 < z < 1.1$, which we split again into five redshift bins. We refer the interested reader to Ferrero et al. (2021) for further details and highlight here only the basic features of the ICE-COLA mocks.

[7] cosmo.nyu.edu/roman/2LPT (Crocce, Pueblas & Scoccimarro 2006; Scoccimarro et al. 2012).
[8] https://wwwmpa.mpa-garching.mpg.de/gadget/ (Springel 2005).
[9] http://popia.ft.uam.es/AHF/ (Knollmann & Knebe 2009).

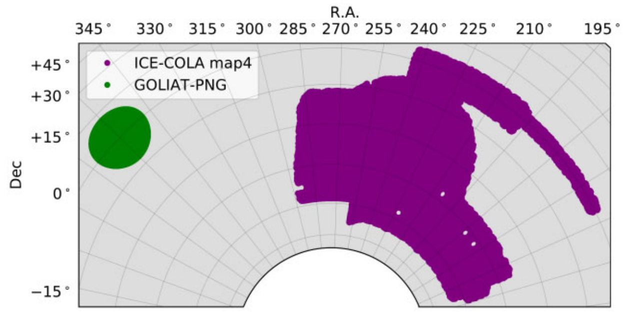

Figure 4. Comparison of the footprint of the used simulations. In purple, we show the mask for one map of the ICE-COLA simulations. In green, we show the mask for the GOLIAT-PNG simulation.

A total number of 488 fast *N*-body simulations of full-sky light cones generated by following the ICE-COLA code Izard, Crocce & Fosalba (2016) are used. This code is based on the COmoving Lagrangian Acceleration (COLA) method, which solves for the evolution of the matter density field using 2LPT combined with a Particle-Mesh (PM). The simulations use 2048^3 particles in a box of the size of 1536 Mpc h^{-1} and assume a cosmology consistent with the best-fitting of the Wilkinson Microwave Anisotropy Probe (WMAP) 5-yr data (Komatsu et al. 2009). This means compatible with a flat ΛCDM model with $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, $\Omega_b = 0.044$, $n_s = 0.95$, $\sigma_8 = 0.8$, $h = 0.7$, and $f_{NL} = 0$. A summary of the cosmological parameters and fiducial values used is presented in the second part of Table 1.

A hybrid halo occupation distribution – halo abundance matching model is used to populate haloes with galaxies. Also, automatic calibration is run to match the basic characteristics of the DES Y3 BAO sample: the observed abundance of galaxies as a function of photometric redshift (Fig. 3), the distribution of photometric redshift errors, and the clustering amplitude on scales smaller than those used for BAO measurements.

Finally, four footprint masks corresponding to the DES Y3 BAO sample are placed on each full-sky light cone simulation to reach the final set of 1952 ICE-COLA mocks. In Fig. 4, we can see the shape of the mask followed by one footprint.

5 ANALYSIS TOOLS

This section presents the statistical tools used to measure the f_{NL} parameter using the theoretical template presented in Section 2.

5.1 ACF measurements

The angular correlations are measured using CUTE (Alonso 2012), which computes the ACF following the Landay–Szalay estimator (Landy & Szalay 1993),

$$w_{\rm obs}(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}, \quad (40)$$

where $DD(\theta)$, $DR(\theta)$, and $RR(\theta)$ are the number counts of pairs of galaxies for the data–data, data–random, and random–random catalogues, respectively. To obtain the random–random pairs, we create random catalogues with 20 times more objects than the simulation sample that follow the angular mask of the simulations for GOLIAT-PNG and ICE-COLA mocks. The random–random pairs are obtained as an output from CUTE.

As mentioned in Section 3.3, one of the key elements in the IC correction is the random-random pairs that account for the survey volume. Because of this, we need to compute at least one $RR(\theta)$ correlation for both GOLIAT-PNG and ICE-COLA going up to the maximum angular separation allowed for each survey mask. That is 22 deg for the GOLIAT-PNG simulations and 88 degrees for ICE-COLA simulations.

5.2 Covariance

Our default set-up for the covariance matrix uses the COSMOLIKE code (Krause & Eifler 2017; Fang, Eifler & Krause 2020a; Fang et al. 2020b) to estimate the covariance analytically. Following Crocce, Cabré & Gaztañaga (2011b), the real space covariance of the ACF $w(\theta)$ at angles θ_i and θ_j is related to the covariance of the angular power spectrum $C(C_\ell, C_{\ell'})$ by

$$C(\theta_i, \theta_j) = \sum_{\ell, \ell'} \frac{(2\ell+1)(2\ell'+1)}{(4\pi)^2} \overline{P_\ell}(\theta_i)\overline{P_{\ell'}}(\theta_j) C(C_\ell, C_{\ell'}), \quad (41)$$

where $\overline{P_\ell}(\theta)$ are the Legendre polynomials averaged over each angular bin and $C(C_\ell, C_{\ell'})$, under the Gaussian approximation, is given by (Crocce et al. 2011b; Krause & Eifler 2017)

$$C(C_\ell, C_{\ell'}) = \frac{2\delta_{\ell\ell'}}{f_{\rm sky}(2\ell+1)} \left(C_{\ell'} + \frac{1}{n_g} \right)^2, \quad (42)$$

where δ is the Kronecker delta function, n_g is the number density of galaxies per steradian, and $f_{\rm sky}$ is the observed sky fraction used to account for partial-sky surveys. We include redshift space distortions through the C_ℓ's of the expression above (42), except when analysing the GOLIAT-PNG mocks, as they do not include that. In addition, following Troxel et al. (2018), we correct the shot-noise contribution to the covariance (the term $\propto 1/n_g$) by considering the effect of the survey geometry on the number of galaxies in each angular bin. We ignore non-Gaussian terms in the covariance estimation for simplicity, following DES Collaboration (2022b), where it was tested that including those terms did not impact the results. See DES Collaboration (2022b) and Ferrero et al. (2021) for the validation of this analytical covariance matrix (with $\theta_{\rm max} = 5$ deg) against two sets of simulations: ICE-COLA and FLASK lognormal mocks (Xavier, Abdalla & Joachimi 2016).

Notice that we do not include f_{NL} in our covariance since it is customary in this kind of analysis to fix the cosmology and then look for deviations. In the case of detection, we should modify the covariance and include the f_{NL} parameter.

We also consider using the ICE-COLA covariance constructed from the mocks, given by

$$C(\theta_i, \theta_j) = \frac{1}{N_m - 1} \sum_{n=1}^{N_m} \left(w^n(\theta_i) - \bar{w}(\theta_i) \right) \left(w^n(\theta_j) - \bar{w}(\theta_j) \right) \quad (43)$$

where N_m is the number of mocks, $w^n(\theta)$ is the ACF for the n-mock, and $\bar{w}(\theta)$ is the mean ACF from the mocks. However, it was shown in Ferrero et al. (2021) that, due to a large number of simulated boxes used to equal the volume of the DES Y3 BAO sample, a replication of halos were produced, introducing a spurious correlation among the measured ACF. This induced a high degree of correlation of non-adjacent redshift bins in the covariance. For this reason, the default set-up of using COSMOLIKE covariance was preferred (DES Collaboration 2022b). As a double check, in Section 7.4, we compare the impact of changing the covariance when measuring f_{NL}.

5.3 Parameter inference

In order to measure the parameters, we perform a Bayesian parameter inference based on the log-likelihood analysis assuming a Gaussian likelihood, as follows:

$$\log(\mathcal{L}(\boldsymbol{p})) \propto -\frac{\chi^2(\boldsymbol{p})}{2} \quad (44)$$

where the χ^2 is given by

$$\chi^2(\boldsymbol{p}) = (M(\boldsymbol{p}) - D)^T C^{-1} (M(\boldsymbol{p}) - D), \quad (45)$$

where $\boldsymbol{p}$ represents the free parameters from our theory we want to estimate, C^{-1} is the inverse of the covariance matrix presented in Section 5.2, and M and D are the theoretical model and the data vector, respectively.

Since the galaxy sample for the simulations is divided into five redshift bins, we perform a joint sampling of the likelihood to consider covariance between bins. The joint data vector $\boldsymbol{D}$ is given by

$$\boldsymbol{D} = [w^1_{\rm obs}(\theta), w^2_{\rm obs}(\theta), w^3_{\rm obs}(\theta), w^4_{\rm obs}(\theta), w^5_{\rm obs}(\theta)], \quad (46)$$

where the superscript represents the redshift bin from which the ACF is obtained. We repeat the same procedure for the theoretical model, where $\boldsymbol{M}(\boldsymbol{p})$ is the theory vector as a function of the free parameters for each redshift bin, as follows:

$$\boldsymbol{M}(\boldsymbol{p}) = [w^1_{\rm th}(\theta, \boldsymbol{p}), w^2_{\rm th}(\theta, \boldsymbol{p}), w^3_{\rm th}(\theta, \boldsymbol{p}), w^4_{\rm th}(\theta, \boldsymbol{p}), w^5_{\rm th}(\theta, \boldsymbol{p})]. \quad (47)$$

We perform an MCMC sampling of the likelihood function using COBAYA (Torrado & Lewis 2021) to estimate the posterior distributions of the free parameters in our pipeline.

Table 1 presents the free parameters considered for our analysis and their respective fiducial values and priors. Depending on the analysis, the IC could be considered as a free parameter (IC-MARG) or fixed to its theoretical value (IC-FIXED) given by equation (39). This will be stated for each test considered. Notice that we are not including other free cosmological parameters in the likelihood, which is customary for this kind of analysis, since adding other cosmological parameters will lose the constraints on $f_{\rm NL}$.

6 TESTS WITH NON-GAUSSIAN MOCKS

This section tests the pipeline over the GOLIAT-PNG simulations with non-Gaussian initial conditions. For these simulations, the theoretical template is obtained from a linear power spectrum without considering BAO damping and without RSD modelling, since the simulations do not include RSD. The goal of this section is twofold: First, we want to recover the fiducial value of $f_{\rm NL}$ for the non-Gaussian simulations. Secondly, we want to highlight the importance of the IC.

6.1 Effect of the IC on GOLIAT-PNG mocks

In Section 3, we presented the IC as one of the key elements that need to be included in the theory. In this section, we show its effect in the simulations with non-Gaussian initial conditions.

In Fig. 5, we compute the IC, as presented in equation (39), for the GOLIAT-PNG simulations but changing the limit angular separation, $\theta_{\rm lim}$, truncating the sum. We use this to test the need to consider the full volume of the survey when computing the IC. As described in Section 4.1, the maximum circular semi-aperture of the GOLIAT-PNG simulation mask is about 11.2 deg, implying that the maximum allowed angular separation is about $\theta_{\rm lim} \sim 22$ deg (vertical grey dotted line in Fig. 5).

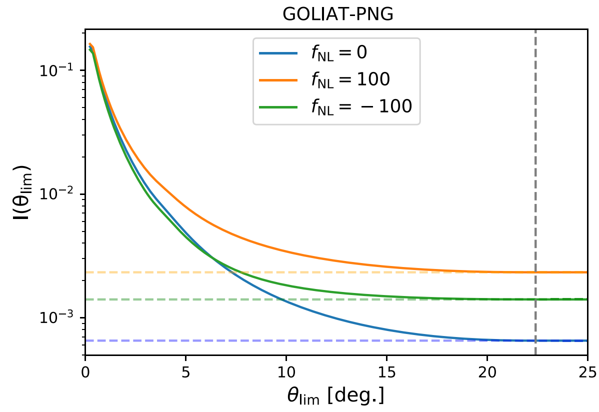

Figure 5. IC as a function of the upper limit angular separation, $\theta_{\rm lim}$, for the GOLIAT-PNG simulations. The blue line is the IC using the theoretical ACF with $f_{\rm NL} = 0$. The same is repeated for the orange and green lines but for the cases of $f_{\rm NL} = 100$ and $f_{\rm NL} = -100$, respectively. The grey dotted line is the limit angular aperture of the angular mask of the simulations.

As expected, given the discussion in Section 3, the IC reaches its value when it is summed up to the maximum angular separation allowed for the simulation mask to consider the whole survey volume. In other words, even though we can compare the theory and the data up to some maximum angular separation $\theta_{\rm max}$, we still need the random–random correlation up to the limit scale of the simulation ($\theta_{\rm lim} \sim 22$ deg). We see that the IC's value does not converge earlier than that. We repeat this conclusion for the ICE-COLA simulations, where the measurements are made up to $\theta_{\rm max} = 20$ deg, but the IC is obtained from random–random pairs measured up to $\theta_{\rm lim} \sim 88$ deg.

From the previous figure, we can also notice the explicit dependence of the IC on $f_{\rm NL}$. For $f_{\rm NL} = 0$, it has a smaller value in comparison with $f_{\rm NL} = 100$ or $f_{\rm NL} = -100$. This supports the previous discussion from Section 3.2 about the importance of the IC when looking for $f_{\rm NL}$.

The previously computed IC can be included in the theory as in equation (37). This is shown in Fig. 6, where we compare the theoretical ACF with and without the IC against the mean of the GOLIAT-PNG mocks. The ACF is shown for the first redshift bin with the errors obtained from the standard deviations of the mocks.

Fig. 6 serves as a visual guide of the IC's effect in the theoretical modelling. The IC correction appears to have an effect that could help avoid biased values for $f_{\rm NL}$. The actual impact of this on the measurement of $f_{\rm NL}$ is the main topic of the following subsection.

6.2 Results for GOLIAT-PNG mocks

We use the parameter inference method, described in Section 5.3, to put constraints on both the linear bias and $f_{\rm NL}$. We construct the data vector for each mock by combining the ACF of each redshift bin for the $f_{\rm NL} = -100$ and $f_{\rm NL} = 100$ simulations. We use the scale configuration given in the first section of Table 4. The scale choice will be justified in the next section when we test the robustness of the pipeline.

Since each mock is independent of the other, we can compute a joint posterior distribution by multiplying the posteriors of $f_{\rm NL}$ and b of each GOLIAT-PNG mock. The advantage of this method is that the joint posterior gives us a good estimate of how biased the best-fitting values of $f_{\rm NL}$ are with respect to the fiducial. We compare fixing the IC, as computed using equation (39), against not using

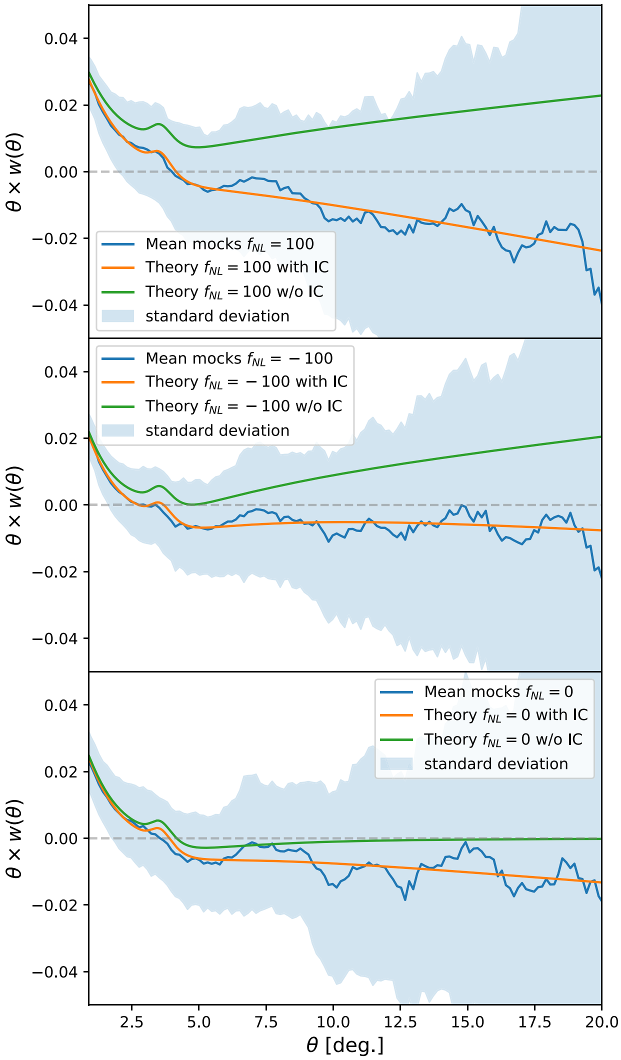

Figure 6. Comparison of the theoretical ACF against the mean GOLIAT-PNG mocks for the first redshift bin ($0.6 < z < 0.7$) and different $f_{\rm NL}$ values. The blue line is the mean of the mocks, and the shaded area is given by its standard deviation. The orange line is the theoretical ACF with the IC. The green line is the theoretical ACF without the IC.

it and against leaving it as a nuisance parameter. The priors for the parameters used in the measurement are in Table 1. For the case of $f_{\rm NL} = 100$ simulations, four mocks were discarded due to incompatibilities in the measurements of $f_{\rm NL}$, giving highly biased values and complicating the computation of the joint posterior.

We present one of the main results of this work in Fig. 7, showing the contours obtained from the joint posterior of all GOLIAT-PNG simulation with $f_{\rm NL} = 100$ and $f_{\rm NL} = -100$. We show that by fixing the IC to the value given by equation (39), we can recover the fiducial values of $f_{\rm NL}$ within 1σ. We also notice that for the case of not using the IC, we obtain very biased values for $f_{\rm NL}$, closer to $f_{\rm NL} = 0$. The figure also shows that when considering the IC as a nuisance parameter and marginalizing it, we also recover the correct values for $f_{\rm NL}$. With the previous results, we prove the importance of the IC.

The summary of contours is presented in Table 2, where we show the measured values of $f_{\rm NL}$ for the two kinds of GOLIAT-PNG simulations. The best-fitting values of $f_{\rm NL}$ are obtained from the maximum of the joint posterior distribution of all mocks, with the errors obtained from the 68 per cent confidence region. We clarify that the uncertainty presented in Table 2 corresponds to the combination of all mocks. This implies that the uncertainty would be ∼16 times larger for a survey with the properties of the GOLIAT-PNG mocks, making the uncertainty and the offset very similar $\Delta f_{\rm NL} \sim \sigma \sim 100$. We also note that the relatively small footprint of GOLIAT-PNG ($\sim 400\text{deg}^2$) makes the effect of the IC stronger. We will reexamine this for a DES-like scenario in Section 7.1.

A natural question appears when we see the results for the case of IC-MARG. Can the marginalized IC case recover the theoretical values given by equation (39)? In Table 3, we compare the IC values for both theoretical and marginalized, along with the 1σ errors for the marginalized case measured over the mean of the mocks. From these results, we can notice two things. First, we found reasonable compatible values for the IC within $\sim 2\sigma$. Secondly, we show that the methods can detect the IC at high significance.

In Fig. 8, we compare the mean of the GOLIAT-PNG $f_{\rm NL} = 100$ mocks versus the theoretical ACF (for IC-FIXED) using the best-fit results with and without the IC for each redshift bin. The figure shows how the IC improves the agreement of the theoretical template and the observed ACF for each redshift bin. Nevertheless, we found no considerable difference in χ^2 of the measurement over the individual mocks when considering or not the IC in the theoretical template. The showed errors, in this case, are obtained from the COSMOLIKE covariance, described in Section 5.2, but divided by the number of mocks, in contrast with the errors presented in Fig. 6.

For the case of NO-IC, we notice that for both simulations, we obtain biased small negative values of $f_{\rm NL}$. As mentioned by the end of Section 2.4, for large negative values of $f_{\rm NL}$ (without considering IC), there is a positive correlation function at large scales (see e.g. middle panel of Fig. 6). Since the measured ACF shows a negative correlation at large scales (due to the observational IC), the model prefers small negative $f_{\rm NL}$ values to compensate for the lack of IC in the theoretical model (see e.g. Fig. 8)

As mentioned in Section 3.2, the effect of the IC is stronger for non-Gaussian simulations due to its dependence on $f_{\rm NL}$. Nevertheless, in the next section, we will show that it can also help avoid slightly biased values of $f_{\rm NL}$ even for simulations with $f_{\rm NL} = 0$, such as the ICE-COLA mocks.

7 DES VALIDATION USING ICE-COLA MOCKS

As mentioned in Section 4, the ICE-COLA mocks are designed to match the DES Y3 BAO sample angular mask and redshift distribution $N(z)$. In this section, we present tests made over the ICE-COLA mocks, assessing their impact on the measurement of the $f_{\rm NL}$ parameter.

We perform four different tests over the ICE-COLA simulations that we briefly summarize as follows:

(i) **Effect of the IC:** Similarly to Section 6, this test double-check the importance of the IC.

(ii) **Best-fitting estimator comparison:** This test will tell us how the value of $f_{\rm NL}$ changes when we consider a different definition for the estimator of the best fit from the posterior distribution.

(iii) **BAO damping versus linear theory:** We will show the impact of considering BAO damping in the theoretical modelling by comparing it with the linear power spectrum.

(iv) **Covariance comparison:** For robustness, we consider different covariances and study their impact on the measurement of $f_{\rm NL}$.

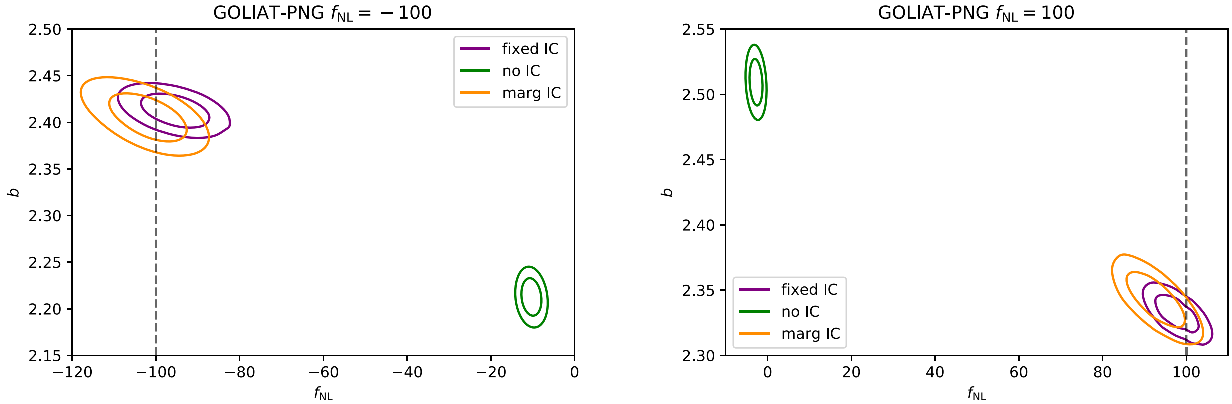

Figure 7. Marginalized one and 2σ contours for f_{NL} and the linear bias b obtained from the joint posterior of the 246 GOLIAT-PNG simulations (of $\sim 400\,\mathrm{deg}^2$ each). Note that the error is expected to be ~16 times larger for a single realization. The left-hand panel is for the $f_{NL} = -100$ simulation, and the right-hand panel is for the $f_{NL} = 100$ simulation. The purple contours are with the IC fixed to its theoretical value given by equation (39). The green contours are without considering any IC correction. The orange contours consider the IC as a nuisance parameter and marginalizing it. The vertical dashed line represents the fiducial value of f_{NL} for each set of simulations.

Table 2. Summary of the results of measuring f_{NL} from both GOLIAT-PNG simulations (of $\sim 400\,\mathrm{deg}^2$ each). The best-fitting values are obtained from the maximum of the joint posterior of the 246 mocks, and the errors are at 1σ. Note that the error is expected to be ~16 times larger for a single realization.

GOLIAT-PNG	Joint posterior
$f_{NL} = 100$	
NO-IC	-2.8 ± 1.0
IC-FIXED	97.4 ± 3.5
IC-MARG	92.2 ± 4.6
$f_{NL} = -100$	
NO-IC	-10.3 ± 1.5
IC-FIXED	-95.2 ± 5.4
IC-MARG	-101.5 ± 6.5

Table 3. Comparison between theoretical IC versus marginalized values for GOLIAT-PNG simulations. IC theory is computed using the theoretical value given by equation (39). IC marginalized are obtained as the mean of the marginalized posterior. The errors are at 1σ on the ensemble average of 246 mocks.

	GOLIAT-PNG	
Redshift bin	IC theory	IC marginalized
$0.6 < z < 0.7$	0.00220	0.00247 ± 0.00013
$0.7 < z < 0.8$	0.00202	0.00208 ± 0.00012
$0.8 < z < 0.9$	0.00188	0.00212 ± 0.00012
$0.9 < z < 1.0$	0.00178	0.00203 ± 0.00011
$1.0 < z < 1.1$	0.00171	0.00184 ± 0.00011

(v) **Scale configuration:** We compare the effect that different scale cuts and theta binning have when estimating f_{NL}.

The fiducial scale configuration for the tests and forecast, along with the optimal f_{NL} best-fitting estimator, are summarized in Table 4. The parameters to analyse are presented in detail in the second section of Table 1. In summary, we consider the linear bias for each redshift bin, the IC as a possible nuisance parameter, and the non-Gaussianity parameter f_{NL}.

For the analysis, we compare two cases: We perform the MCMC sampling for each mock separately and the mean of the mocks. A summary of the results of this section is presented in Table 5. The first column presents the mean of the best-fitting value of f_{NL}, $\langle \hat{f}_{NL} \rangle$, for the ICE-COLA mocks, obtained from the mean of the best-fitting estimator of each mock, $\hat{f}_{NL}$. The second column presents the overall standard deviation in f_{NL}, obtained from the standard deviation of $\hat{f}_{NL}$ coming from each mock. The third column is the mean of the 1σ error obtained from the f_{NL} posterior of each mock. The fourth column is the value of f_{NL} obtained from fitting the theory over the mean of the mocks. The errors over the mean mocks are from the 68 per cent confidence level of the marginalized posterior distribution. It is worth remembering that the ICE-COLA mocks have $f_{NL} = 0$ as an initial condition.

The results from this section are presented in Fig. 9, where for each test, we show the histogram of the best-fitting values, $\hat{f}_{NL}$, from each mock. We also show the mean of the histogram, $\langle \hat{f}_{NL} \rangle$, for each test.

After the tests, we forecast the accuracy in the measurement of the local primordial non-Gaussianity parameter f_{NL} using the ACF with IC over DES Y3 data. We would be able to obtain an error of $\sigma(f_{NL}) = 31$ if the measurement is performed over the DES Y3 BAO sample, as we will see by the end of the section.

7.1 Effect of the IC on ICE-COLA mocks

Here we show the effect of the IC over the ICE-COLA simulations. We compare the effect of the IC for three different cases:

(i) Without using any IC correction (no IC).

(ii) Fixing the IC to the value obtained using equation(39), following the discussion from Section 6.1 (fixed IC).

(iii) Considering the IC as a nuisance parameter and marginalizing over it (marg IC).

The results are presented in the top panel of Fig. 9 and summarized in the first part of Table 5. From the first column of the table, we notice that not using the IC gives a biased value of f_{NL}, with a deviation of $\Delta f_{NL} \sim 7$ from the fiducial value of the simulation. We also notice that leaving the IC as a free nuisance parameter gives slightly larger errors for f_{NL}. Finally, we show that fixing the IC to the value given

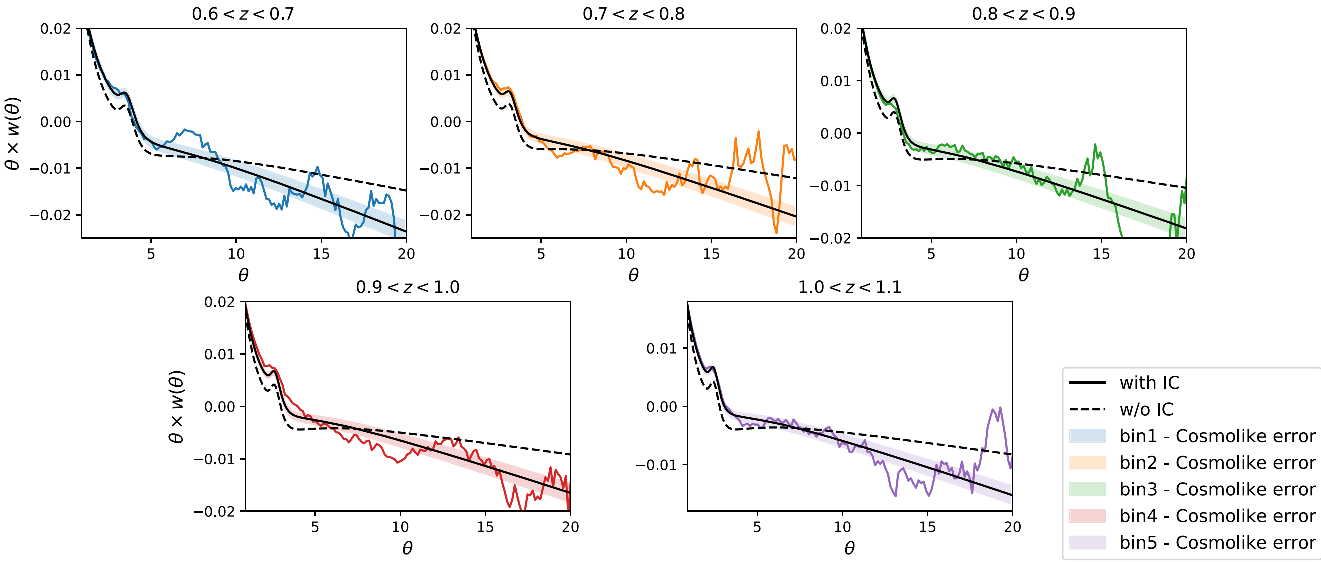

Figure 8. Comparison between the theoretical angular correlation versus the mean ACF of the GOLIAT-PNG mocks with $f_{\rm NL} = 100$, for each redshift bin. The solid-coloured lines are the mean of the ACF from the mocks. The shaded areas are obtained from the diagonal of the reduced theoretical covariance. The solid black lines are theoretical ACF with IC, where $f_{\rm NL}$ and b_g are obtained from the mean of the joint posterior distribution presented in purple in Fig. 7. The dashed black lines are the theoretical ACF without IC and $f_{\rm NL}$ and b_g obtained from the mean of the joint posterior distribution presented in the green lines of Fig. 7.

Table 4. Fiducial configuration of the ACF for both the GOLIAT-PNG and ICE-COLA mocks.

	$\theta_{\rm min}$	$\theta_{\rm max}$	$\Delta\theta$	$f_{\rm NL}$ estimator
GOLIAT-PNG	1.0 deg	20 deg	0.15 deg	Max of marg. posterior
ICE-COLA	1.0 deg	20 deg	0.4 deg	Max of marg. posterior

by equation (39) gives almost no bias in $f_{\rm NL}$, recovering the fiducial value of $f_{\rm NL} = 0$ with high accuracy. Similar to the conclusion from Section 6, the IC helps us to avoid biased values of $f_{\rm NL}$. Although this effect was stronger for non-Gaussian mocks, for the case of $f_{\rm NL} = 0$, we can still notice a difference when measuring $f_{\rm NL}$.

The mild deviation on $f_{\rm NL}$ due to not including the IC on ICE-COLA mocks ($\Delta f_{\rm NL} \sim 7$) opposes the significant bias coming from non-Gaussian mocks ($\Delta f_{\rm NL} \sim 100$). Part of this difference is expected to come from a stronger IC effect on smaller mocks (GOLIAT-PNG), but another important effect comes from the IC being stronger mocks with PNG, as we discussed in much detail in Section 3. In order to separate those effects, we now run our fit on a theory-data vector generated for $f_{\rm NL} = 100$ in a DES-like scenario, including the IC and based on the ICE-COLA cosmology.

From the posterior distribution of Fig. 10, we found, as expected, that we recover the fiducial value, $f_{\rm NL} = 99 \pm 16$, for the case of fixed-IC. Whereas for the case of ignoring the IC, we found $f_{\rm NL} = 76 \pm 13$. The deviation of $\Delta f_{\rm NL} \sim 23$ corresponds to a 1.8σ bias in the value of $f_{\rm NL}$ in a non-Gaussian (DES-Y3-like) scenario. The bias also translates into a mild deviation of $\Delta\chi^2 \sim 2$ in favour of using the IC in the theoretical model. Even though the bias on $f_{\rm NL}$ is not as strong as for the GOLIAT-PNG simulations, we still see a more biased value than the case of $f_{\rm NL} = 0$ simulations. We also observed a higher error in $f_{\rm NL}$ when using $f_{\rm NL} = 0$ compared to $f_{\rm NL} = 100$ as fiducial values, with a difference of approximately $\Delta\sigma(f_{\rm NL}) \sim 15$. This difference appears because the increment in the correlation function is stronger as $f_{\rm NL}$ increases.

Table 5. Summary of measuring $f_{\rm NL}$ from the ICE-COLA mocks. The first column is the overall best fit of $f_{\rm NL}$ obtained as the mean of $f_{\rm NL}$ from each mock. The second column is the error in $f_{\rm NL}$ from the standard deviation of every histogram. The third column is the mean of 1σ error from the $f_{\rm NL}$ posterior of each mock. The fourth column is the value of $f_{\rm NL}$ when fitting over the mean of the mocks. The errors are obtained at the 68 per cent confidence level of the posterior. In bold, we highlight the fiducial configuration that will be used for the forecast.

ICE-COLA	$\langle \hat{f}_{\rm NL} \rangle$	std($\hat{f}_{\rm NL}$)	$\langle \sigma(\hat{f}_{\rm NL}) \rangle$	mean of mocks
NO-IC	− 7.4	26.6	22.1	-12 ± 22
IC-FIXED	0.1	31	24.8	-4.5 ± 24
IC-MARG	4.2	35	29	-3 ± 27
Mean posterior	− 6.6	30.9	–	–
Max posterior	0.1	31	–	–
Min χ^2	0.06	31.1	–	–
Damping	0.1	31	24.8	-4.5 ± 24
Linear	2.4	30.5	24.6	-2.2 ± 23
COSMOLIKE cov.	0.1	31	24.8	-4.5 ± 24
ICE-COLA cov.	− 0.3	29.6	25.4	-9 ± 28
$w(\theta)[\Delta\theta = 0.1]$	− 2.2	32.2	25.4	-7.5 ± 24
$w(\theta)[\Delta\theta = 0.2]$	− 1.7	32.4	25.5	-6.5 ± 24
$w(\theta)[\Delta\theta = 0.3]$	− 0.9	31.8	25.1	-5.5 ± 24
$\boldsymbol{w(\theta)[\Delta\theta = 0.4]}$	0.1	31	24.8	-4.5 ± 24
$w(\theta)[\theta_{\rm max} = 5]$	3.6	35.1	30	-1.7 ± 27
$w(\theta)[\theta_{\rm max} = 10]$	0.6	33.4	26.7	-3.7 ± 26
$w(\theta)[\theta_{\rm max} = 15]$	− 0.08	32.4	25.4	-6.2 ± 24
$\boldsymbol{w(\theta)[\theta_{\rm max} = 20]}$	0.1	31	24.8	-4.5 ± 24

The previous test can be repeated for a theory-data vector with $f_{\rm NL} = -100$, where we found $f_{\rm NL} = -95 \pm 28$ for the fixed-IC case and $f_{\rm NL} = -80 \pm 23$ for the no-IC case. In this case, the bias is less significant: $\Delta f_{\rm NL} \sim 20$, approximately 1σ. Hence, even for a

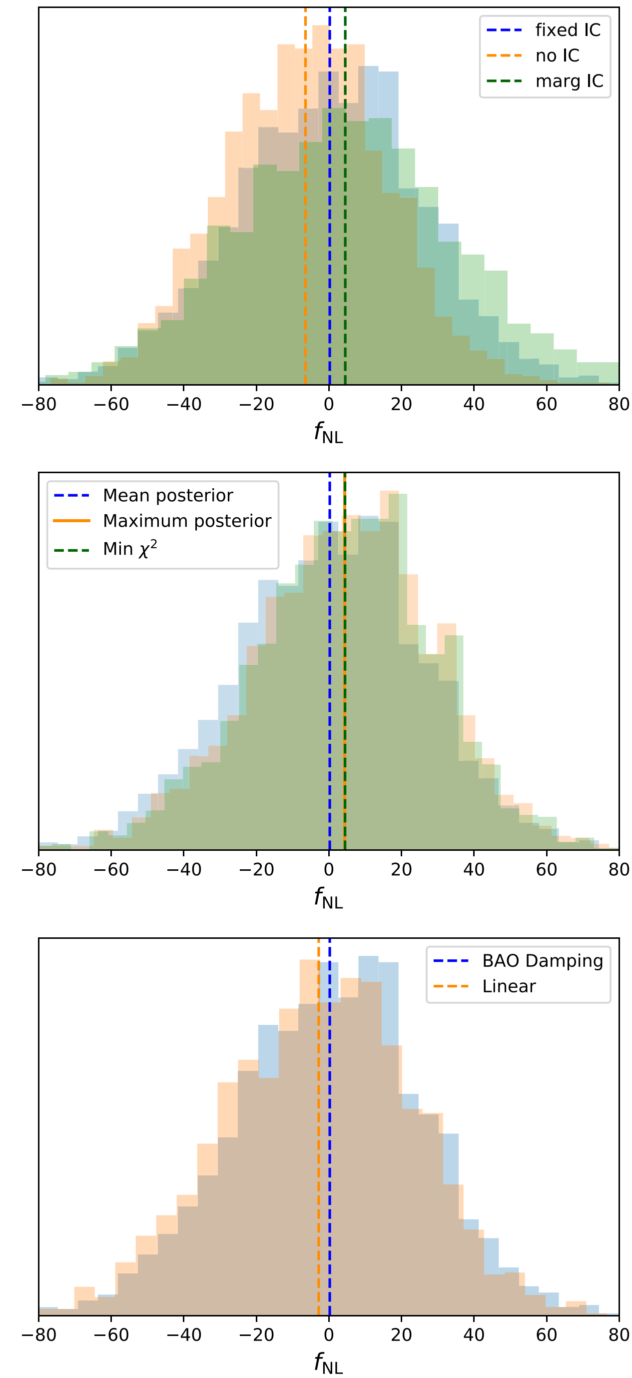

Figure 9. Histograms of the $f_{\rm NL}$ measurement over the 1952 ICE-COLA mocks comparing the different tests. The vertical dotted lines represent the mean of the histograms. **Top panel**: Effect of the IC. The blue is with fixing the IC as in equation (37) (fixed IC), the yellow is without using the IC (no IC), and the green is the IC as a nuisance parameter (marg IC). The vertical dotted lines represent the mean of the histograms. **Middle panel:** Best-fitting estimator comparison. The blue is the mean of the posterior as the best fit, the yellow uses the maximum of the posterior (MAP), and the green uses the minimum of the χ^2. **Bottom panel:** Raw linear theory versus BAO damping comparison. The blue includes BAO damping in the template, and the orange uses the linear theory.

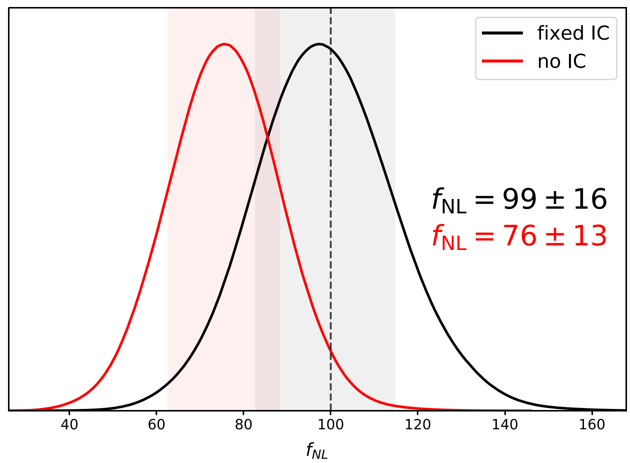

Figure 10. Marginalized $f_{\rm NL}$ posteriors resulting from fitting our model to a theory-data vector with $f_{\rm NL}$ = 100 with IC and a DES-like set-up (∼ 4100deg^2, see Table 1). We use both a model with IC (black line) and without IC (red line), finding consistency for the former and a 1.8σ bias for the latter. The shaded areas represent the $f_{\rm NL}$ marginalized errors at 68 per cent C.L.

large DES-like area, the bias on $f_{\rm NL}$ when ignoring the IC becomes significant if the data we are fitting contains PNG. Similar to the conclusion from Section 6, the results highlight the importance of the IC when dealing with primordial non-Gaussianity.

7.2 Best-fitting estimator comparison

We compare different ways to extract the best-fitting estimator $\hat{f}_{\rm NL}$ from the marginalized $f_{\rm NL}$ posterior distribution, that is, using different central tendency estimators. We show the differences between using the mean of the marginalized posterior, the maximum of the marginalized posterior (MP), or the minimum of the χ^2.

The comparison of the histograms is presented in the second panel of Fig. 9. The summary of the results from this test is also presented in Table 5. From the table, we can see that we found no considerable differences in using the maximum of the posterior distribution and the minimum of the χ^2. Furthermore, we notice an improvement when we use the maximum of the posterior, against the mean of the posterior, as an estimator of the central value for $f_{\rm NL}$, where we found almost no bias. In the end, the maximum of the posterior was preferred.

7.3 Linear theory versus BAO damping

As mentioned in the theoretical modelling, we focused on the damping model because of its improvement when fitting the BAO peak. One open question is whether we need to consider such precision in the template when measuring $f_{\rm NL}$.

To address the previous question, we compare the $f_{\rm NL}$ measurement from an ACF with a BAO damping model against using the ACF from a linear power spectrum. Both ACFs are computed using the fiducial configuration.

We summarize the results in Table 5. We found that using a linear power spectrum introduces a small bias compared to including the BAO damping in the power spectrum.

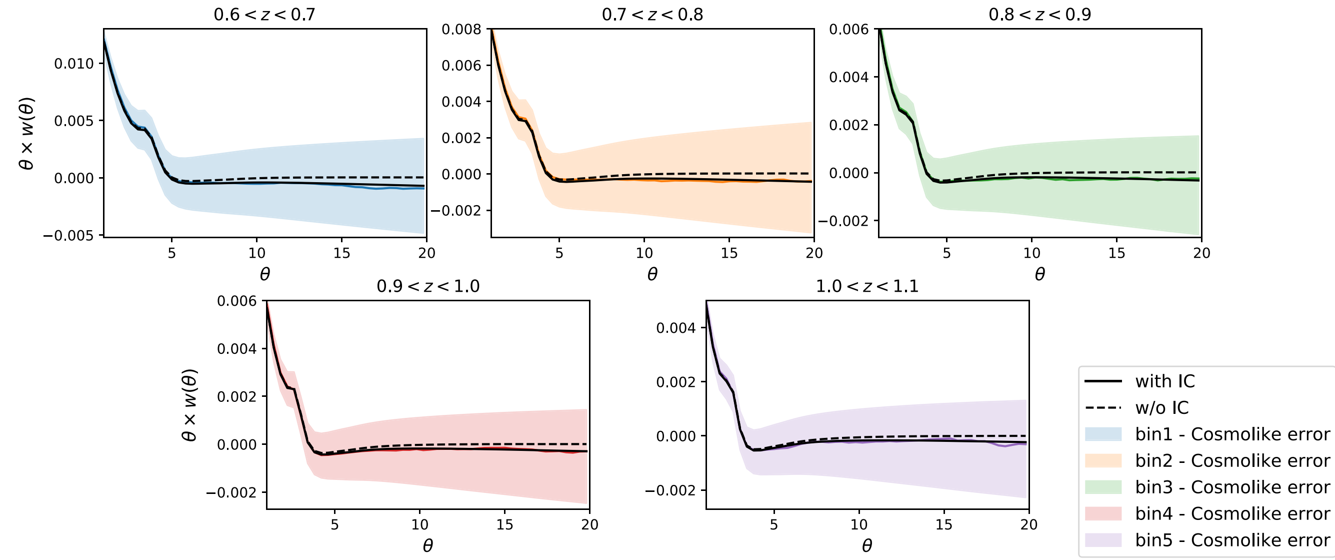

Figure 11. Comparison between the theoretical ACF versus the mean ACF of the ICE-COLA mocks for each redshift bin. The solid black lines are theoretical ACF computed for $\langle \hat{f}_{\rm NL} \rangle$ and b_g obtained using the optimal fiducial configuration and fixing the IC. The black dashed lines are the theoretical ACF without IC. The solid-coloured lines are the mean of the ACF from the mocks. The shaded areas are errors obtained from the COSMOLIKE covariance.

7.4 Covariance comparison

In Section 5.2, we mentioned that the default covariance matrix used is COSMOLIKE since the ICE-COLA presented a spurious correlation between non-adjacent redshift bins. In this section, we compare the effect of different covariance in the $f_{\rm NL}$ measurements. We compare the COSMOLIKE covariance versus the covariance obtained from the ICE-COLA mocks.

The results are presented in Table 5, where we show that the measurement of $f_{\rm NL}$ is robust against changes in the covariance.

7.5 Scale configuration

In this section, we discuss the impact of different scale configurations on the measurement of $f_{\rm NL}$. We compute the theory and the data vector from each ICE-COLA mock considering a combination of the following scales:

(i) $\theta_{\rm max} = [5, 10, 15, 20]$ deg.
(ii) $\Delta\theta = [0.1, 0.2, 0.3, 0.4]$ deg.

We summarize the extracted information in the last two sections of Table 5. For this study, we limited ourselves to a maximum angular separation of 20 deg because we consider that controlling the LSS systematics up to these scales will already be very challenging. Note that the fiducial maximum angular scale for the BAO measurement was 5 deg (DES Collaboration 2022b).

From the results, we can notice two effects. First, the measurements of $f_{\rm NL}$ seem to be robust against the change in $\Delta\theta$, introducing small changes in both the mean and its error. The second effect appears when we go to larger values of $\theta_{\rm max}$, where there is an ~ 11 per cent improvement in the constraints when going up to $\theta_{\rm max} = 20$. This improvement is expected since most of the $f_{\rm NL}$ effect comes from large scales.

From the results of this section, we have three main conclusions: First, we can improve the accuracy of $f_{\rm NL}$ by using the IC. Not including it is the main source of bias in our measurement, introducing deviations of $\Delta f_{\rm NL} \sim 7$ to the fiducial value. In the second place, we can improve the precision on $f_{\rm NL}$ constraints by ~ 11 per cent when going to angular scales of $\theta_{\rm max} = 20$. Thirdly, our analysis is robust against changes in the type of covariance, the inclusion of BAO damping, and changes in the scale binning, where we found almost no deviations in the precision and accuracy of $f_{\rm NL}$. These conclusions allow us to define the fiducial configuration highlighted in Table 4.

Using the fiducial configuration, in Fig. 11, we show the ACF for the best-fitting values compared against the mean of the ICE-COLA mocks for each redshift bin with and without the IC, fixed to the value given by equation (39). From Fig. 11, we can notice the importance of the IC when comparing the model with the simulations improving its matching, especially at large scales, and therefore, improving the accuracy of $f_{\rm NL}$.

After the tests from this section, we conclude that a reliable forecast is $\sigma(f_{\rm NL}) = 31$ for the DES Y3 BAO sample after marginalizing the linear bias and fixing the other cosmological parameters. The forecast is also done using the fiducial configuration from Table 4.

8 CONCLUSIONS

We have presented a methodology to constrain $f_{\rm NL}$ using the two-point ACF with scale-dependent bias. Primordial non-Gaussianity modifies the linear bias relation between dark matter overdensities and galaxies by including a scale dependence that depends on the $f_{\rm NL}$ parameter. The scale dependency is later introduced in the power spectrum and transferred to the ACF. It is worth noticing that there are differences in the effect of the scale-dependent bias; for the power spectrum, the effect is more localized, whereas for the (angular) correlation function, it is more extended over a range of scales.

We remarked on the importance of the IC condition, an observational constraint that appears due to the limited volume observed by surveys and the fact that we estimated the mean number density from them. This condition is essential because of the $f_{\rm NL}$ effect in the two-point correlation at large scales and the divergent behaviour of the power spectrum at $k \rightarrow 0$ (see equation 8). We impose the IC condition on our theoretical model and show that it can be corrected by a constant.

We tested the model with the IC correction against the GOLIAT-PNG simulations with non-Gaussian initial conditions. We showed how the IC is a crucial element in avoiding biased $f_{\rm NL}$ values. We showed that ignoring the IC gives very biased PNG constraints, $f_{\rm NL} = -2.8 \pm 1.0$ ($f_{\rm NL} = -10.3 \pm 1.5$), whereas we recover the fiducial value $f_{\rm NL} = 100$ ($f_{\rm NL} = -100$), within 1σ, when correcting for the IC: $f_{\rm NL} = 97.4 \pm 3.5$ ($f_{\rm NL} = -95.2 \pm 5.4$). We confirmed the importance of the IC for simulations with $f_{\rm NL} = 100$ and $f_{\rm NL} = -100$.

We used the ICE-COLA mocks to validate and test the robustness of the pipeline against different analysis choices when measuring $f_{\rm NL}$. We showed that fixing the IC (equation 39) improves the accuracy in the value of $f_{\rm NL}$, correcting for a $\Delta f_{\rm NL} \sim 7$ deviation with respect to the fiducial value when not including it. Furthermore, we showed that going to large angular scales of $\theta_{\rm max} = 20$ improves the $f_{\rm NL}$ precision by ~ 11 per cent. In addition, we showed that not including the BAO damping can introduce a slight bias of $\Delta f_{\rm NL} \sim 2$. Also, our results prove to be robust against changes in the choice of covariance matrices and the choice of angular binning. Using a theory-data vector with $f_{\rm NL} = 100$ ($f_{\rm NL} = -100$) with IC based on ICE-COLA cosmology, area, and n(z), we also checked the importance of the IC when having the realistic case of a DES-Y3-like survey. We found a $\Delta f_{\rm NL} \sim 23$ ($\Delta f_{\rm NL} \sim 15$) deviation when not using the IC in our theoretical modelling.

One of the main conclusions of this paper is that when ignoring the IC in a PNG analysis, we always find a bias on the recovered $f_{\rm NL}$. This bias is strongest for a small survey and a *true universe* with PNG (GOLIAT-PNG : $\Delta f_{\rm NL} \sim 100 \sim \sigma$). For a large survey like DES, we still find a significant bias on $f_{\rm NL}$ for a *true universe* with PNG ($\Delta f_{\rm NL} \sim 20 \sim 1 - 2\sigma$, for $f_{\rm NL}^{\rm true} = 100$). Whereas the bias on $f_{\rm NL}$ is mild for a large survey ($\sim$4100 deg^2) and a Gaussian *true universe* ($\Delta f_{\rm NL} \sim 7 \sim 0.3\sigma$).

We expect our analysis to be the first step into constraining $f_{\rm NL}$ with the Dark Energy Survey photometric data, where we forecast a measurement of $f_{\rm NL}$ within $\sigma(f_{\rm NL}) = 31$ when measured against the DES Y3 BAO sample. This prospect is comparable with the current constraints coming from spectroscopic surveys, being $\sigma(f_{\rm NL}) \sim 21$ (Mueller et al. 2021) the latest one to date.

Future plans include mitigation of LSS systematics following up on Carnero Rosell et al. (2022) and Rodríguez-Monroy et al. (2022) with a particular focus on very large scales (see e.g. Rezaie et al. 2021), which is crucial as systematic errors due to survey properties can lead to spurious PNG signal (Ross et al. 2013; Mueller et al. 2021). Given this, we plan to conduct a full battery of robustness tests while blinded to the $f_{\rm NL}$ value, following the standard DES policy (DES Collaboration 2022b). Additionally, performing PNG analysis can also be understood as a strong validation exercise of the galaxy clustering systematics, given the sensitivity of this probe to them. Note also that future photometric surveys are expected to break the barrier of $\sigma(f_{\rm NL}) = 1$ (de Putter & Doré 2017), key to the inflationary models, and this work is a necessary step towards that goal.

ACKNOWLEDGEMENTS

We thank the anonymous reviewer for insightful comments and suggestions. The work of WR is funded by a fellowship from 'La Caixa' Foundation (ID 100010434) with fellowship code LCF/BQ/DI18/11660034 and the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 713673. SA is currently supported by the Spanish Agencia Estatal de Investigacion through the grant 'IFT Centro de Excelencia Severo Ochoa by CEX2020-001007-S', he was also supported by the MICUES project, funded by the EU H2020 Marie Skłodowska-Curie Actions grant agreement no. 713366 (InterTalentum Fellowship UAM). WR, SA, and JGB acknowledge support from the Research Project PGC2018-094773-B-C32 and the Centro de Excelencia Severo Ochoa Program CEX2020-001007-S. WR, SA, and JGB acknowledge the use of the Hydra cluster at IFT.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NSF's NOIRLab, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

This study is based in part on observations at Cerro Tololo Inter-American Observatory at NSF's NOIRLab (NOIRLab Prop. ID 2012B-0001; PI: J. Frieman), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under grant numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under grants ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA programme of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) do e-Universo (CNPq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Alonso D., 2012, preprint (arXiv:1210.1833)
Angulo R. E., Pontzen A., 2016, MNRAS, 462, L1
Avila S., Adame A. G., 2023, MNRAS, 519, 3706
Barreira A., 2020, J. Cosmol. Astropart. Phys., 2020, 031
Beutler F. et al., 2014, MNRAS, 443, 1065
Blas D., Garny M., Ivanov M. M., Sibiryakov S., 2016, J. Cosmol. Astropart. Phys., 2016, 028
Breton M.-A., de la Torre S., 2021, A&A, 646, A40
Byrnes C. T., Choi K.-Y., 2010, Adv. Astron., 2010, 724525
Cabass G., Ivanov M. M., Philcox O. H. E., Simonović M., Zaldarriaga M., 2022, Phys. Rev. D, 106, 043506
Carnero Rosell A. et al., 2022, MNRAS, 509, 778
Castorina E. et al., 2019, J. Cosmol. Astropart. Phys., 2019, 010
Castorina E., Feng Y., Seljak U., Villaescusa-Navarro F., 2018, Phys. Rev. Lett., 121, 101301
Chan K. C. et al., 2018, MNRAS, 480, 3031
Chan K. C., Hamaus N., Biagetti M., 2019, Phys. Rev. D, 99, 121304
Crocce M., Cabré A., Gaztañaga E., 2011a, MNRAS, 414, 329
Crocce M., Cabré A., Gaztañaga E., 2011b, MNRAS, 414, 329
Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369
Dalal N., Doré O., Huterer D., Shirokov A., 2008, Phys. Rev. D, 77, 123514
de Mattia A., Ruhlmann-Kleider V., 2019, J. Cosmol. Astropart. Phys., 2019, 036
de Putter R., Doré O., 2017, Phys. Rev. D, 95, 123513
DES Collaboration, 2018, Phys. Rev. D, 98, 043526
DES Collaboration, 2019, MNRAS, 483, 4866
DES Collaboration, 2021, ApJS, 255, 20
DES Collaboration, 2022a, Phys. Rev. D, 105, 023520
DES Collaboration, 2022b, Phys. Rev. D, 105, 043512
Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1
Fang X., Eifler T., Krause E., 2020a, MNRAS, 497, 2699
Fang X., Krause E., Eifler T., MacCrann N., 2020b, J. Cosmol. Astropart. Phys., 2020, 010
Ferrero I. et al., 2021, A&A, 656, A106
Fillmore J. A., Goldreich P., 1984, ApJ, 281, 1
Giannantonio T., Ross A. J., Percival W. J., Crittenden R., Bacher D., Kilbinger M., Nichol R., Weller J., 2014, Phys. Rev. D, 89, 023511
Gil-Marín H., Percival W. J., Verde L., Brownstein J. R., Chuang C.-H., Kitaura F.-S., Rodríguez-Torres S. A., Olmstead M. D., 2017, MNRAS, 465, 1757
Groth E. J., Peebles P. J. E., 1977, ApJ, 217, 385
Ho S. et al., 2015, J. Cosmol. Astropart. Phys., 2015, 040
Howlett C., Lewis A., Hall A., Challinor A., 2012, J. Cosmol. Astropart. Phys., 2012, 027
Ivanov M. M., Sibiryakov S., 2018, J. Cosmol. Astropart. Phys., 2018, 053
Izard A., Crocce M., Fosalba P., 2016, MNRAS, 459, 2327
Jeong D., Komatsu E., 2009, ApJ, 703, 1230
Knollmann S. R., Knebe A., 2009, ApJS, 182, 608
Komatsu E. et al., 2009, ApJS, 180, 330
Komatsu E., Spergel D. N., 2001, Phys. Rev. D, 63, 063002
Krause E., Eifler T., 2017, MNRAS, 470, 2100
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Leistedt B., Peiris H. V., Roth N., 2014, Phys. Rev. Lett., 113, 221301
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Linder E. V., 2005, Phys. Rev. D, 72, 043529
LSST Science Collaboration 2009, preprint (arXiv:0912.0201)
Matarrese S., Verde L., 2008, ApJ, 677, L77
Moradinezhad Dizgah A., Biagetti M., Sefusatti E., Desjacques V., Noreña J., 2021, J. Cosmol. Astropart. Phys., 2021, 015
Mueller E.-M. et al., 2021, preprint (arXiv:2106.13725)
Mueller E.-M., Percival W. J., Ruggeri R., 2019, MNRAS, 485, 4160
Pajer E., Schmidt F., Zaldarriaga M., 2013, Phys. Rev. D, 88, 083502
Peacock J. A., 1999, Cosmological Physics. Cambridge University Press, Cambridge
Peacock J. A., Nicholson D., 1991, MNRAS, 253, 307
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton University Press, New Jersey
Planck CollaborationIX, 2020, A&A, 641, A9
Porredon A. et al., 2021, Phys. Rev. D, 106, 103530
Rezaie M. et al., 2021, MNRAS, 506, 3439
Rodríguez-Monroy M. et al., 2022, MNRAS, 511, 2665
Ross A. J. et al., 2013, MNRAS, 428, 1116
Scoccimarro R., Hui L., Manera M., Chan K. C., 2012, Phys. Rev. D, 85, 083002
Slosar A., Hirata C., Seljak U., Ho S., Padmanabhan N., 2008, J. Cosmol. Astropart. Phys., 2008, 031
Springel V., 2005, MNRAS, 364, 1105
Sugiyama N. S., Saito S., Beutler F., Seo H.-J., 2019, MNRAS, 484, 364
Tasinato G., Tellarini M., Ross A. J., Wands D., 2014, J. Cosmol. Astropart. Phys., 2014, 032
Torrado J., Lewis A., 2021, J. Cosmol. Astropart. Phys., 2021, 057
Troxel M. A. et al., 2018, MNRAS, 479, 4998
Xavier H. S., Abdalla F. B., Joachimi B., 2016, MNRAS, 459, 3693

APPENDIX: THE ANALYTIC CORRELATION FUNCTION, THE IC AND $f_{\rm NL}$

In order to gain some insight about the divergent behaviour mentioned in Section 3.2 due to the theoretical IC with $f_{\rm NL}$, in this appendix, we compute the explicit dependence on the large scales of the IC condition for the two-point correlation function.

Let us start by considering the primordial power spectrum $P_\Phi(k) = A\,k^{n_s}$, which can be used to define the linear matter power spectrum by considering a simplified transfer function (Peacock 1999),

$$P_m(k) = P_\Phi(k)\,T^2(k) = \frac{A\,k^{n_s}}{(1+k^2/k_{\rm eq}^2)^2}, \tag{A1}$$

where $k_{\rm eq}$ is the wavenumber at matter-radiation equality. As previously seen in Section 2, from the matter power spectrum, we can compute the multipole expansion of the two-point correlation function using equation(13).

For simplicity, we focus on the monopole. It is possible to compute the 2PCF for $n = 1$ and $k_{\rm eq} = 1$ as

$$\xi_0(r) = \frac{1}{4\pi^2 r}\,\left(g(r) + r\,g'(r)\right), \tag{A2}$$

$$g(r) = \cosh\,r\,{\rm shi}\,r - \sinh\,r\,{\rm chi}\,r, \tag{A3}$$

where shi() and chi() are the Sinh- and Cosh-Integral functions. This shows that the 2PCF can be computed analytically for this power spectrum.

The next step is to show analytically how the 2PCF changes if we include a scale-dependent bias and use the simplified matter power spectrum. Let us start by recalling the expression of the power spectrum with scale-dependent bias,

$$P_g(k) = b(k)^2 P_m(k), \tag{A4}$$

$$b(k) = b_g + \Delta b(k, z). \tag{A5}$$

We find terms that are independent, linear, and quadratic in $f_{\rm NL}$. This implies that the computation of the 2PCF involves three integrals over the wavenumbers. The term *independent of* $f_{\rm NL}$ just gives something proportional to $b_g^2\,\xi_0(r)$.

The *linear term in* $f_{\rm NL}$ is more interesting. This component of 2PCF is proportional to,

$$\int_0^\infty {\rm d}k k^2 \frac{P_m(k)}{k^2\, T(k)} j_0(kr) = \frac{g(r)}{r}, \tag{A6}$$

which is finite for large values of r.

The *quadratic term in* $f_{\rm NL}$ logarithmically diverges as $k_{\rm min} \rightarrow 0$, this can be seen as follows:

$$\int_{k_{\rm min}}^\infty k^2 \frac{P_m(k)}{k^4\, T(k)^2}\, j_0(k\, r){\rm d}k = j_0(k_{\rm min} r) - {\rm ci}(k_{\rm min} r) \tag{A7}$$

$$\rightarrow 1 - \gamma - \ln(k_{\rm min} r) + \frac{1}{12} k_{\rm min}^2 r^2, \tag{A8}$$

where ci is the Cosine Integral function.

Now that we have computed the 2PCF for the simplified power spectrum with scale-dependent bias, we can analyse how the theoretical IC condition, given by equation(35), behaves at large scales.

From the previous computation can be seen that the IC condition explicitly vanishes for the term independent of $f_{\rm NL}$,

$$\int_0^\infty {\rm d}r r^2 \xi_0(r) = 0. \tag{A9}$$

The linear term, given by equation (A6), is linearly divergent for a given large scale $r_{\rm max}$. This can be seen as follows:

$$\int_0^{r_{\rm max}} {\rm d}r r^2 \frac{g(r)}{r} = r_{\rm max}(1 + g'(r_{\rm max})) - g(r_{\rm max}) \rightarrow r_{\rm max}\,, \tag{A10}$$

This implies that there will be a linear term in $f_{\rm NL}$ proportional to $f_{\rm NL} r_{\rm max}$.

Now if we compute the IC for the quadratic term in $f_{\rm NL}$, given by equation (A7), we find that it is proportional to $f_{\rm NL}^2 k_{\rm eq}^3/k_{\rm min}^3$.

Therefore, from this calculation, we conclude that the IC has a term linear in $f_{\rm NL}$, which diverges with $r_{\rm max}$, and a quadratic term in $f_{\rm NL}^2$ which is proportional to $k_{\rm min}^{-3}$. This implies that, even for infinite volume surveys, we need to correct the two-point correlation function with PNG with the IC because it can bias the $f_{\rm NL}$ results.

[1] *Instituto de Física Teorica UAM-CSIC, c/Nicolás Cabrera 13-15, E-28049 Madrid, Spain*
[2] *Departamento de Física Teórica, Universidad Autónoma de Madrid, E-28049 Madrid, Spain*
[3] *Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*
[4] *Department of Physics, The Ohio State University, Columbus, OH 43210, USA*
[5] *Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK*
[6] *Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, NO-0315 Oslo, Norway*
[7] *School of Physics and Astronomy, Sun Yat-sen University, 2 Daxue Road, Tangjia, Zhuhai 519082, China*
[8] *CSST Science Center for the Guangdong-Hongkong-Macau Greater Bay Area, SYSU, 519082, China*
[9] *ICTP South American Institute for Fundamental Research, Instituto de Física Teórica, Universidade Estadual Paulista, 01140-070 São Paulo, Brazil*
[10] *Instituto de Física Teórica, Universidade Estadual Paulista, 01140-070 São Paulo, Brazil*
[11] *Laboratório Interinstitucional de e-Astronomia – LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil*
[12] *Instituto de Astrofisica de Canarias, E-38205 La Laguna, Tenerife, Spain*
[13] *Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[14] *Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[15] *Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*
[16] *Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA*
[17] *Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr, Pasadena, CA 91109, USA*
[18] *Department of Astronomy, University of California, Berkeley, 501 Campbell Hall, Berkeley, CA 94720, USA*
[19] *Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA*

This paper has been typeset from a TEX/LATEX file prepared by the author.